

Ancestral Processes with Selection*

Stephen M. Krone[†]

University of Idaho

and

Claudia Neuhauser[‡]

University of Minnesota

Received November 12, 1995

In this paper, we show how to construct the genealogy of a sample of genes for a large class of models with selection and mutation. Each gene corresponds to a single locus at which there is no recombination. The genealogy of the sample is embedded in a graph which we call the *ancestral selection graph*. This graph contains all the information about the ancestry; it is the analogue of Kingman's coalescent process which arises in the case with no selection. The ancestral selection graph can be easily simulated and we outline an algorithm for simulating samples. The main goal is to analyze the ancestral selection graph and to compare it to Kingman's coalescent process. In the case of no mutation, we find that the distribution of the time to the most recent common ancestor does not depend on the selection coefficient and hence is the same as in the neutral case. When the mutation rate is positive, we give a procedure for computing the probability that two individuals in a sample are identical by descent and the Laplace transform of the time to the most recent common ancestor of a sample of two individuals; we evaluate the first two terms of their respective power series in terms of the selection coefficient. The probability of identity by descent depends on both the selection coefficient and the mutation rate and is different from the analogous expression in the neutral case. The Laplace transform does not have a linear correction term in the selection coefficient. We also provide a recursion formula that can be used to approximate the probability of a given sample by simulating backwards along the sample paths of the ancestral selection graph, a technique developed by Griffiths and Tavaré (1994). © 1997 Academic Press

1. INTRODUCTION

This paper is concerned with the description of the genealogical process for standard models in population

* This work was begun while the first author was at the University of Wisconsin, Madison. This work was completed while the second author visited the Department of Ecology and Evolutionary Biology, Princeton University, and the Institute for Advanced Study, School of Mathematics, Princeton, NJ 08540.

[†] Partially supported by NSF Grant 9353149 (Presidential Faculty Fellowship held by R. Pemantle).

[‡] Alfred P. Sloan Research Fellow. Partially supported by the National Science Foundation under grant DMS-9403644.

genetics in which both mutation and selection are present, such as the discrete-time Wright–Fisher model (i.e., non-overlapping generations) or its continuous-time cousin the Moran model (i.e., overlapping generations). In these models, a haploid population of size N undergoes random mating subject to the influence of mutation and selection. Kingman (1982a, b) studied models without selection and found that a fruitful way to investigate such populations is to examine the ancestral history of a given sample. This leads to the genealogical process under neutrality, known as Kingman's coalescent. Finding the corresponding process when selection is

present, has been an open problem. Here we describe the genealogical processes for standard population genetics models with mutation and selection, and thus provide the analogue of Kingman's coalescent for models with selection. Our derivation of the genealogical process is based on the continuous-time Moran model (Moran 1958, 1962) since the continuous-time model is easier to analyze than the discrete-time Wright–Fisher model. We wish to emphasize that for large populations the discrete-time and the continuous-time models behave essentially the same and hence our results will *not* depend on the particular choice of model.

With this in mind, we now consider a haploid population of N individuals that evolves according to a Moran model with selection and mutation. (Haploid means that each individual has only one copy of each chromosome.) For each individual, we focus on a single locus of a given chromosome and limit the discussion to two alleles (types), A_1 and A_2 . One of the alleles may have a selective advantage. Furthermore, we assume that the mutation is symmetric. (We will comment on more general settings in Section 6.) The Moran model with selection and mutation is a continuous-time Markov process in which individuals produce offspring, one at a time, at rates depending on their types. The type of the offspring will be chosen according to the mutation process. The offspring will then replace an individual chosen at random from the population. (The offspring may replace its own parent. In some formulations, the offspring is not allowed to replace its parent. Either choice yields the same results; there is only a slight change in time scale.) The replaced individual is removed from the population, thus keeping the population size constant. We assume that an individual of type A_i , $i = 1, 2$, reproduces at rate λ_i . The offspring will have the same type as the parent with probability $1 - u_N$ and will have the other type with probability u_N . The parameter $u_N \in [0, 1]$ denotes the probability of mutation per gene per birth event. Throughout the paper we assume that $\lambda_2 \geq \lambda_1$ to indicate that type A_2 has a selective advantage (or at least no disadvantage). We set

$$\lambda_2 = \lambda_1(1 + s_N) \tag{1.1}$$

with $s_N \geq 0$.

The state of the population at time t can be represented as a continuous-time Markov chain $Z(t) = (Z_1(t), Z_2(t))$, where $Z_i(t)$, $i = 1, 2$, denotes the number of individuals of type A_i in the population at time t . (Of course, $Z_1(t) + Z_2(t) = N$.) If $Z_1(t) = j$, $j = 0, 1, 2, \dots, N$, the following transitions occur:

$$j \rightarrow \begin{cases} j+1 & \text{at rate } \lambda_1 j \frac{(N-j)}{N} (1-u_N) \\ & + \lambda_2 (N-j) \frac{(N-j)}{N} u_N \\ j-1 & \text{at rate } \lambda_2 (N-j) \frac{j}{N} (1-u_N) \\ & + \lambda_1 j \frac{j}{N} u_N \end{cases} \tag{1.2}$$

Thus, $Z_1(t)$ is a birth-death process with nonlinear birth and death rates which can be analyzed using standard Markov chain theory. It is customary to study the limiting ($N \rightarrow \infty$) diffusion approximation of this model. To carry out this limit, one needs to rescale the parameters of the process. We set

$$\lambda_2 = \lambda_1(1 + s_N), \quad \text{with } \lambda_1 = \frac{N}{2} \tag{1.3}$$

and assume that

$$Ns_N \rightarrow \sigma \quad \text{and} \quad Nu_N \rightarrow \theta \quad \text{as } N \rightarrow \infty, \tag{1.4}$$

with $\sigma \geq 0$, $\theta \geq 0$, and

$$|Ns_N - \sigma| \leq CN^{-\gamma} \quad \text{and} \quad |Nu_N - \theta| \leq CN^{-\gamma} \tag{1.5}$$

for some $\gamma > 0$ and $C < \infty$. (For example, $s_N = \sigma/N$ and $u_N = \theta/N$ satisfy (1.4) and (1.5) trivially.) We denote the fraction of genes of type A_i at time t in the limiting process (as $N \rightarrow \infty$) by $X_i(t)$ so that $X(t) \equiv (X_1(t), X_2(t))$ is a random variable in the set $\mathcal{A}_2 = \{(x_1, x_2) : x_1, x_2 \geq 0, x_1 + x_2 = 1\}$. The diffusion process $X_1(t) \in [0, 1]$ has drift $a(x) = -(\sigma/2)x(1-x) + (\theta/2)(1-2x)$ and diffusion parameter $b(x) = x(1-x)$. This process has a unique stationary distribution with density given by

$$h(x) = Kx^{\theta-1}(1-x)^{\theta-1}e^{-\sigma x}, \quad x \in [0, 1], \tag{1.6}$$

where K is a normalizing constant. (It is readily verified that $h(x)$ satisfies

$$0 = \frac{1}{2} \frac{\partial^2}{\partial x^2} [b(x)h(x)] - \frac{\partial}{\partial x} [a(x)h(x)],$$

and hence $h(x)$ in (1.6) is a stationary distribution.) The density in (1.6) is a special case of Wright's formula (Wright, 1949, p. 383) which was later verified by Kimura (1956).

Instead of studying the evolution of the population as a whole, one is frequently interested in properties of finite samples from the population. Due to kinship, the sample points are not independent. Understanding the genealogy of a sample is therefore of great importance if, for example, one wishes to estimate population parameters. In the neutral case, that is, when $s_N = 0$ (or, equivalently, $\lambda_1 = \lambda_2$), the genealogy of a sample of size n can be described by the n -coalescent which was introduced in the seminal work of Kingman (1982a, b). This process has been the object of numerous research papers and has led to many beautiful results; see, e.g., Tavaré (1984), Donnelly (1984), Watterson (1984), Griffiths (1989, 1991), Griffiths and Tavaré (1994). This list is, of course, very incomplete and we refer the reader to the forthcoming book by Donnelly and Tavaré (1997) for much more on this subject. By contrast, almost nothing is known about genealogical structure in population models with selection; the only results are those by Kaplan, Darden and Hudson (1988) and Hudson and Kaplan (1988). (The latter paper includes recombination in addition to selection. We will comment on their results in Section 6.)

Our goal is to introduce and investigate the analogue of Kingman's coalescent in the selective case; that is, we will study in detail the genealogical relationships between a finite number of genes sampled from a large population ($N \rightarrow \infty$) that evolves according to the Moran model described in (1.2). (Extensions to other population models are discussed in Section 6.)

This paper is organized as follows. In Section 2 we will give an alternative description of $Z(t)$ in terms of the biased voter model on a complete graph. Taking the diffusive limit ((1.3)–(1.5)) and using duality, we can construct an object which is analogous to the coalescent, and which we call the *ancestral selection graph*. In Section 3 we will analyze the ancestral selection graph, with and without mutation, and give an algorithm for simulating it. In Section 4 we will discuss a procedure to compute the probability that two individuals in a sample are identical by descent. This can be computed easily in the neutral case. We will see that with selection it is much more complicated and we will only be able to obtain the first two terms in its power series about $\sigma = 0$. The same method can also be used to obtain the power series about $\sigma = 0$ of the Laplace transform of the time to the most recent common ancestor for a sample of two individuals. This is also done in Section 4. In Section 5, we derive a recursion formula that can be used to simulate backwards along the sample paths of the ancestral selection graph. This method was developed by Griffiths and Tavaré (1994) in the neutral case where they use Markov chain Monte Carlo techniques to provide a very efficient

way to estimate the probability of a particular sample configuration. Section 6 is primarily devoted to extending our results to other population models. The main results of this paper are Theorems 2.5, 3.2, 3.15, 4.19, 4.26 and 5.2. Some of the mathematical details are found in the appendices.

2. THE BIASED VOTER MODEL, DUALITY, AND THE ANCESTRAL SELECTION GRAPH

In this section we describe an alternative formulation of the Moran model in terms of the so-called *biased voter model with mutation on a complete graph*. We begin with a description of its dynamics.

Let $I = \{1, 2, \dots, N\}$ denote the set of sites in a complete graph. (In a complete graph, each site has all other sites as its neighbors, so that there is no spatial structure.) The biased voter model with mutation on a complete graph is a continuous-time Markov process whose state at time t is denoted by $\eta_t: I \rightarrow \{1, 2\}$. So η_t assigns a value 1 or 2 to each site in I . If, for $x \in I$, $\eta_t(x) = 1$ (respectively, 2), then we say that x is occupied by an individual of type A_1 (respectively, A_2) at time t . The process $\{\eta_t; t \geq 0\}$ evolves according to the following rules.

- (i) For $x = 1, 2, \dots, N$, if $\eta_t(x) = 1$ (respectively, 2), the individual at x produces offspring at rate λ_1 (respectively, λ_2).
- (ii) The offspring is of the same type as the parent with probability $1 - u_N$ and of the other type with probability u_N .
- (iii) At the time of the birth event, one of the N sites is chosen at random and the individual at that site is replaced by the offspring. (The offspring is allowed to replace its own parent.)

We assume $\lambda_2 = \lambda_1(1 + s_N)$, with $s_N \geq 0$, and $u_N \in [0, 1]$. It is clear that this process is equivalent to the Moran model with population size N . In particular, the number of sites in state 1 satisfies (1.2). The case $\lambda_1 = \lambda_2$ is called the voter model with mutation on a complete graph and was used by Donnelly (1984) to investigate the neutral Moran model. The voter model (without mutation) on \mathbb{Z}^d was introduced independently by Holley and Liggett (1975) and by Clifford and Sudbury (1973). The biased voter model (without mutation) on \mathbb{Z}^d , the d -dimensional integer lattice, was introduced by Harris (1976) and studied by Schwartz (1977) and by Bramson and Griffeath (1980, 1981). (In the process

on \mathbb{Z}^d , each site has only a finite number of neighbors which introduces a strong spatial dependence and makes the analysis much harder.)

Our process can be constructed using a “graphical representation.” This simple and useful technique was introduced by Harris (1972) and is commonly used in interacting particle systems. The idea is to construct the process with a collection of independent Poisson processes by drawing several types of arrows on the space-time coordinate system $I \times [0, \infty)$; cf. Figure 1. We will refer to this graph as the *percolation diagram*. These arrows will indicate where and when the offspring is produced and where it is sent to. We begin by attaching arrows to each time line $\{x\} \times [0, \infty)$ at the times of arrivals of a Poisson process that describes the birth process. Note that the birth rate depends on the type of the individual; we wish to construct the process without making reference to the initial configuration at time 0 (or to the current configuration). This can be achieved by running the Poisson process at the larger rate (i.e., at rate λ_2) and using a thinning procedure to produce a Poisson process which will give arrivals at the smaller rate. If the individual at x is of type A_2 , it will produce offspring every time it encounters an arrival in the Poisson process run at rate λ_2 ; if the individual at x is of type A_1 , it will only produce offspring at times when it encounters arrivals in the thinned out Poisson process. According to (iii), we also need to select a site for the offspring. Since the site is chosen uniformly among the N sites, we can define the birth process as follows: For each ordered pair $(x, y) \in I^2$, we run a Poisson process at rate λ_2/N to indicate that a birth from x to y occurs if the individual at x is of type A_2 . To obtain the arrival process for individuals of type A_1 , we select a random variable which is uniformly distributed over $(0, 1)$ at each arrival point (independently of all other arrivals). If this random variable is $\geq \lambda_1/\lambda_2$, we discard the arrival point. Thus the thinned out process consists of those arrivals for which the uniform variable associated with it is $< \lambda_1/\lambda_2$. (It is an elementary fact that this produces a Poisson process run at rate $\lambda_2 \lambda_1/\lambda_2 = \lambda_1$.) To keep track of these events on the percolation diagram, we will use two types of arrows as defined below. Mutation events can be easily incorporated as shown below.

Motivated by this description, we define the graphical representation rigorously in the following way. For each $x, y \in I$, let $\{W_n^{x,y} : n \geq 1\}$ denote the times of arrivals in a Poisson process with rate λ_2/N . Let $\{U_n^{x,y} : n \geq 1\}$ and $\{V_n^{x,y} : n \geq 1\}$ be sequences of independent, uniformly distributed random variables in $(0, 1)$. We assume that all of these sequences are independent. At times $W_n^{x,y}$ we draw an arrow from x to y to indicate a birth of an off-

spring at x which is then sent to y . If $V_n^{x,y} < \lambda_1/\lambda_2$, we put a δ at the tip of the arrow, otherwise we label the arrow with a 2 (and do not put a δ at the tip of the arrow). That is, into a site y , we have δ -arrows entering at rate λ_1 and 2-arrows at rate $\lambda_2 - \lambda_1$. One thinks of the type of an individual at time 0 spreading up the graph (in the direction of increasing time) and along certain arrows. The rule will then be that 2's can give birth through both types of arrows, but 1's can only give birth through δ -arrows. (In other words, the random variables $V_n^{x,y}$, $n \geq 1$, are used to thin out the rate λ_2/N Poisson process.) The “ δ ” at the tip of the unlabelled arrow indicates the death of the individual at y due to the landing of an offspring from x . The lack of a δ on the tip of the labelled arrows means that a death only occurs when a 2 at x attempts a birth onto y . The process $\{U_n^{x,y} : n \geq 1\}$ is used for the mutation process. If $U_n^{x,y} < u_N$, a mutation occurs. We indicate mutation events by solid dots on the arrows. We wish to point out that the type of the parent determines whether or not its offspring can replace the individual at the landing site of the offspring. In particular, this means that if a mutation occurs at a 2-arrow, the offspring of an A_2 parent will replace the individual at the landing site even though the offspring is of type A_1 ; but the offspring of an A_1 parent (which is of type A_2 if a mutation occurs) cannot replace the individual at the landing site. Tracing the effects of the arrows as time increases, it should be clear that this graphical representation produces the biased voter model on I , and hence gives a construction of the Moran model.

A realization of the percolation diagram in the case $N=9$ can be seen in Figure 1. If initially the set of 1's is $\{4, 8\}$, then at time t , the set of 1's is $\{2, 3, 5, 8, 9\}$. The paths of the 1's are indicated by thick lines.

The advantage of using the graphical representation lies in the fact that by reversing time, we can follow the ancestral history of individuals at sites in a finite set and thus determine their types. The resulting process is called the *dual* or *ancestral process*. Harris (1976) established the fact that the dual process of a biased voter model on \mathbb{Z}^d is a coalescing, branching random walk on \mathbb{Z}^d . Replacing \mathbb{Z}^d and its neighborhood structure with the complete graph, we arrive at the same conclusion: the dual process is again a coalescing, branching random walk, but this time on the complete graph. The random walk chooses which of the N sites to jump to according to a uniform distribution, i.e., probability $1/N$ for each site. The dual process for a voter model ($\lambda_1 = \lambda_2$) is simply a coalescing random walk (that is, without branching); this was exploited by Donnelly (1984) in the analysis of the Moran model in the neutral case.

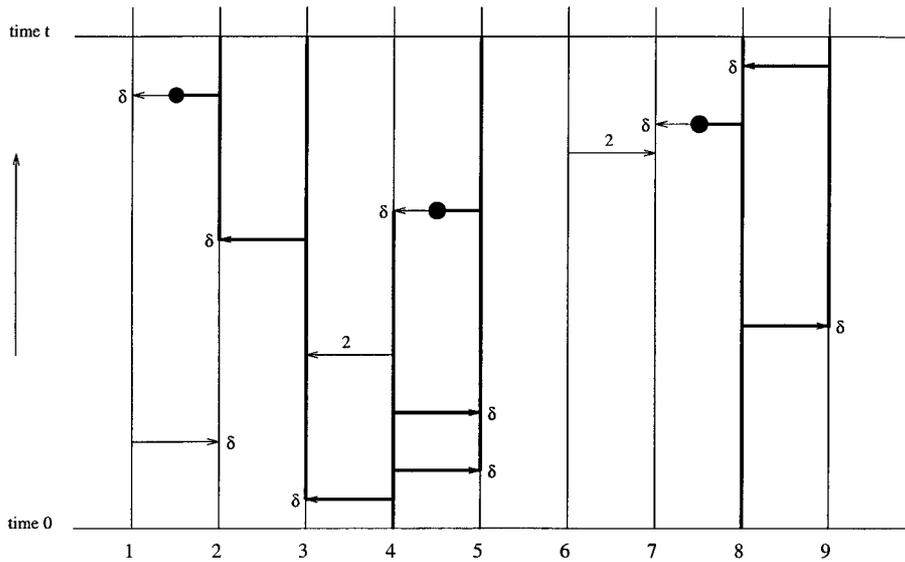


FIG. 1. The graphical representation for the biased voter model for the case $N=9$.

The rules for the dual process are best explained using an example. A realization of the dual process, which was obtained from Figure 1 by simply reversing time and the direction of the arrows in the percolation diagram in Figure 1, can be seen in Figure 2. In particular, time 0 is now at the top of the graph; we refer to this backward-running time as *dual time* to distinguish it from the (forward-running) time in Figure 1. (We will omit the word “dual” if it is clear from the context.) This graph allows us to determine the ancestral history of each individual and consequently allows us to determine their

types if we know the types of the individuals in the population at some time back in the past. We explain this using a sample of individuals located at sites 1, 3 and 5 at dual time 0 in Figure 2. We assume that we know the types of the individuals at dual time t .

Following the time line of the individual located at site 1 at dual time 0 (starting at the top of the graph and going down the vertical line), the individual encounters a “ δ ” at dual time t_1 . This means that, regardless of the type of our individual, it was produced and sent to this site from site 2 at dual time t_1 , as indicated by the arrow;

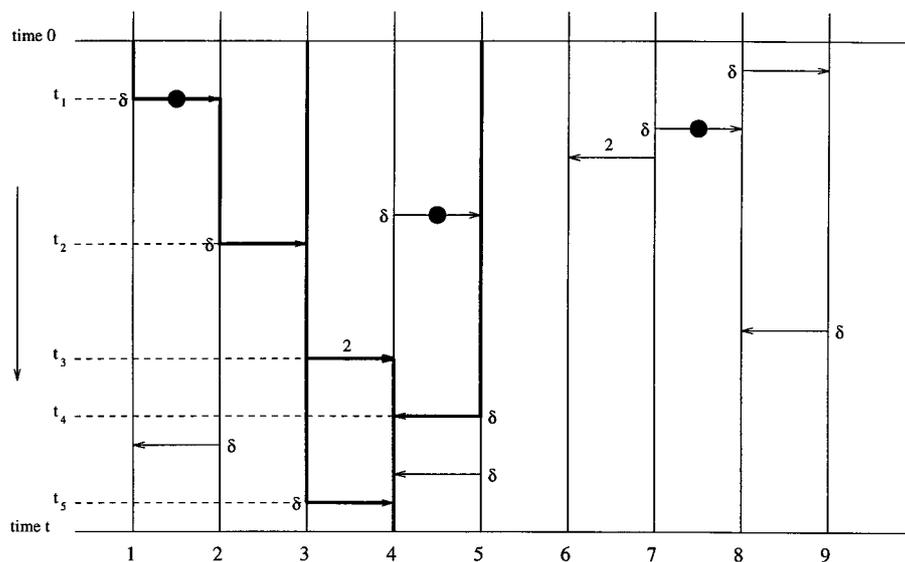


FIG. 2. The graphical representation for the dual process of the biased voter model.

that is, its parent is located at site 2. Following now the parent, we find that regardless of its type, it produces at time t_2 and its parent is located at site 3. The individual site at site 3 encounters a 2-arrow at dual time t_3 . This arrow could have been used only by an individual of type A_2 . At this point, we do not know the type of the individual at site 3 and hence we do not know whether its parent is located at site 4 or whether its path continues on site 3. We thus keep track of both possibilities, and this introduces a branching event in the ancestral graph. The two possibilities are as follows: (1) The individual at site 4 at dual time t_3 is of type A_2 and hence the arrow was used to produce an offspring and the ancestral path of the individual at site 3 continues on site 4. (2) The individual at site 4 at dual time t_3 is of type A_1 and hence the arrow was not used and the ancestral path continues on site 3. Following the first possibility, the ancestral path remains at site 4 until dual time t . The second possibility leads to site 4 as well (at dual time t_5). Since we assumed that we know the types of the individuals at dual time t , we can thus determine the type of the individual at site 1 at dual time 0 by *only* using the type of the individual at site 4 at dual time t and the ancestral paths we constructed for the individual at site 1 at dual time 0. Furthermore, this allows us to establish which of the paths were actually used in the ancestral history of the individual under consideration.

In a similar way, we can construct the ancestral paths for the individuals located at sites 3 and 5 at dual time 0. We wish to point out that at dual time t_2 , the ancestral path of the individual starting at site 3 at dual time 0 coalesces with the ancestral path of the individual starting at site 1 at dual time 0; that is, at dual time t_2 , the two individuals have a common ancestor. Likewise, coalescing events occur at dual times t_4 and t_5 .

We have now constructed the ancestral process of a sample of individuals located at sites 1, 3, and 5 at dual time 0. The paths consist of the thickened lines in Figure 2. The dual or ancestral process we constructed in Figure 2 is typical in the sense that (i) the dual process has a coalescing/branching structure (in fact, the dual process is a coalescing/branching random walk on the complete graph I since every time a jump to a different site occurs, this site is chosen at random from among all sites in the set I), and (ii) starting from a finite subset, the size of the dual process will reach 1 eventually (a fact which will be proved later). Once the dual process reaches size 1, it is enough to know the allele frequencies of the population at this time to reconstruct the types of the individuals by using only the paths in the ancestral process and the type of the common ancestor.

We can summarize the dynamics of the dual process as follows: When the path encounters a “ δ ” it follows the arrow; when it encounters a 2-arrow, it branches. The path that follows the 2-arrow will be called the *incoming path*, the other path will be called the *continuing path*. When two paths hit, they coalesce. Mutation events do not have an effect when determining the paths of the dual process; they do, however, have an effect when going forward in time to determine the true genealogical relationships between sampled individuals, as we will explain later.

To get the analogue of the coalescent, we rescale time and the parameters as in Section 1. Recall,

$$\lambda_1 = \frac{N}{2}, \quad \lambda_2 = \lambda_1(1 + s_N), \quad s_N \geq 0,$$

with (2.1)

$$Ns_N \rightarrow \sigma \quad \text{and} \quad Nu_N \rightarrow \theta \quad \text{as } N \rightarrow \infty,$$

together with the rates of convergence in (1.5). Note that $\lambda_2 - \lambda_1 \rightarrow \sigma/2$ as $N \rightarrow \infty$.

Since there is no spatial structure in the process, one only needs to keep track of branching, coalescing, and mutation events in the dual process to obtain information on ancestral relationships. The exact locations of the dual particles are not important. As we will see, in the limit as $N \rightarrow \infty$, the dual process can be reduced to a simple branching/coalescing graph with mutation events along the branches of its graph. It is customary to draw this “ancestral process” such that time 0 (the present time) is on the bottom of the graph (opposite to the direction in Figure 2). An example of such a graph can be seen in Figure 3. The graph in Figure 3 corresponds to the dual graph in Figure 2. Note that the random walk structure of the percolation diagram has been removed; i.e., the random walk steps are no longer visible. This is due to the absence of spatial structure in the population; the only random walk steps that really matter are the ones that cause coalescing or branching. In the following, when we talk about going backwards on the graph, we mean that we move up in the graph in Figure 3. Likewise, coalescing refers to the coalescing of two branches when going up the graph, and branching refers to branching when going up the graph. In the case of a branching event, we call the branch that corresponds to the incoming path (due to a 2-arrow in the percolation diagram) the *incoming branch*; the other branch is called the *continuing branch*. (In Figure 3, the incoming branch is the branch that takes a turn.)

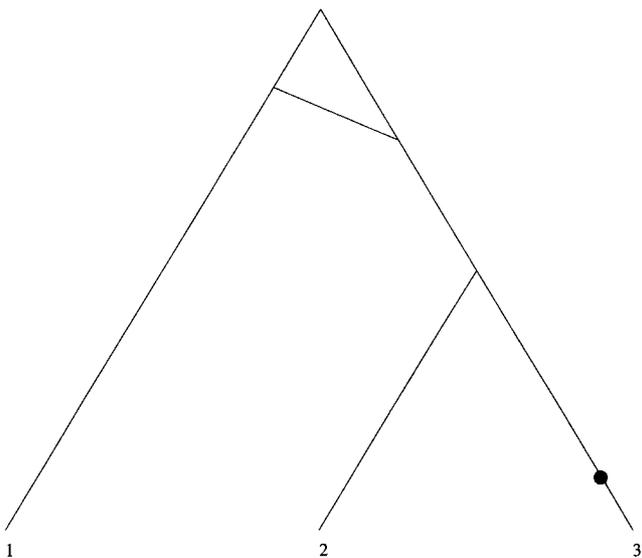


FIG. 3. The ancestral selection graph.

We begin by analyzing how the dual process evolves when starting with a sample of size n at time 0 (the present) taken from a population of size N that evolves according to the Moran model defined above with $u_N = 0$. We denote the dual process by $\{\mathcal{G}_n^N(t): t \geq 0\}$ and refer to its elements as particles and their locations as sites. (Note that we choose not to include the mutation process in the definition of \mathcal{G}_n^N but rather superimpose it afterwards on \mathcal{G}_n^N .)

The dynamics of $\mathcal{G}_n^N(t)$ follow easily from the graphical representation. As already mentioned, since there is no spatial structure in the process, one can ignore the event when a particle in the dual process crosses a δ -arrow and lands either on a site not present in the dual process or on itself. To describe the other possible events, assume that there are k particles in the dual process. We say that a *coalescing event* occurs when a particle crosses a δ -arrow and lands on a particle which is already contained in the dual process, but different from itself. This happens at rate

$$\lambda_1 k \frac{k-1}{N} = \frac{N}{2} k \frac{k-1}{N} = \binom{k}{2}. \quad (2.2)$$

We say that a *branching event* occurs when a particle crosses a 2-arrow. This happens at rate

$$(\lambda_2 - \lambda_1)k = \frac{N}{2} s_N k. \quad (2.3)$$

The original particle continues along the old path (continuing path) and the “offspring” (i.e., the new particle

that arose from the branching) follows the 2-arrow (incoming path). The offspring then either lands on a site which is not contained in the dual process or lands on a site which is already contained in the dual process; the former event occurs with probability $(N-k)/N$, the latter event with probability k/N . If the offspring lands on a site which is already contained in the dual process, we say a *collision* has occurred. We then call the new particle *fictitious* and let it evolve, independently of all other particles, according to the same rules; that is, we define an independent copy of the graphical representation which we use for its movement. (A similar approach was used in Durrett and Neuhauser (1994) for a different dual process.) Offspring of fictitious particles are again fictitious. They use the same graphical representation as their parent particles unless they branch onto a site which is already occupied by a (fictitious or non-fictitious) particle contained in the dual process. We stipulate that fictitious particles can coalesce with both fictitious and non-fictitious particles. If a fictitious particle coalesces with a non-fictitious particle, the resulting (non-fictitious) particle continues on the original graphical representation. If a fictitious particle coalesces with another fictitious particle, the resulting (fictitious) particle continues on the older of the two graphical representations. We will see later that in the limit $N \rightarrow \infty$, collisions can be ignored.

We describe the evolution of the dual process inductively. Let $\mathcal{A}_n^N(t) = \mathcal{A}_n^{N,1}(t) \cup \mathcal{A}_n^{N,2}(t)$ denote the set of particles that are present in the dual at time $t \geq 0$. The set $\mathcal{A}_n^{N,1}(t)$ contains all the non-fictitious particles, whereas $\mathcal{A}_n^{N,2}(t)$ contains the fictitious particles. At time 0, we set $\mathcal{A}_n^{N,1}(0) = \{1, 2, \dots, n\}$ and $\mathcal{A}_n^{N,2}(0) = \emptyset$. Let R_1^N be the smallest value of t at which either a coalescing or branching event occurs. We define a label (β_1^N, γ_1^N) to keep track of what happened at time R_1^N . We set $\beta_1^N = 0$ and $\gamma_1^N = (i, j)$, $i < j$, if a coalescing event occurred at time R_1^N in which particles i and j coalesced. The resulting particle will be labelled i . In this case, $\mathcal{A}_n^{N,1}(R_1^N) = \{1, 2, \dots, n\} \setminus \{j\}$. We set $\beta_1^N = j$ and $\gamma_1^N = (0, 0)$ if a branching event occurred at time R_1^N in which particle j branched; its offspring is labelled $n+1$, regardless of whether it is fictitious or not. If the offspring is non-fictitious, $\mathcal{A}_n^{N,1}(R_1^N) = \{1, 2, \dots, n+1\}$ and $\mathcal{A}_n^{N,2}(t) = \emptyset$; if the offspring is fictitious, $\mathcal{A}_n^{N,1}(R_1^N) = \{1, 2, \dots, n\}$ and $\mathcal{A}_n^{N,2} = \{n+1\}$. We can now pass to the inductive step. Suppose we have defined the process up to time R_m^N with $m \geq 1$. We wait until the next time, R_{m+1}^N , at which either a coalescing or branching event occurs. We define a label $(\beta_{m+1}^N, \gamma_{m+1}^N)$ in exactly the same way as when $m = 1$ and treat fictitious and non-fictitious particles as before. The resulting particle of a coalescing event will be given the

smaller of the two labels of the coalescing particles. The new particle of a branching event will be called $n + m$.

Mutation events occur with probability u_N , $0 \leq u_N \leq 1$, on both types of arrows. Since arrows enter a site at rate λ_2 , mutations along a path occur at rate $\lambda_2 u_N = N(1 + s_N)u_N/2$.

The dual process $\{\mathcal{G}_n^N(t): t \geq 0\}$ consists thus of the three components $\{\mathcal{A}_n^N(t): t \geq 0\}$, $\{R_m^N: m \geq 1\}$, and $\{(\beta_m^N, \gamma_m^N): m \geq 1\}$. On this dual process, we superimpose the mutation process, which we denote by $\{\mathcal{Y}_n^N(t): t \geq 0\}$.

It is convenient to stop the process at the first time at which the size of the dual process reaches 1, since the type of the particle present at that time determines the types in the sample. We call the particle at that time the *ultimate ancestor* and denote the corresponding time by T_{UA}^N .

We are ready to prepare for the main result of this section which says that, in the limit as $N \rightarrow \infty$, the process $\{\mathcal{G}_n^N(t): 0 \leq t \leq T_{UA}^N\}$ converges in distribution to a well-defined process $\{\mathcal{G}_n(t): 0 \leq t \leq T_{UA}\}$ which we call the *ancestral selection graph*. The ancestral selection graph $\{\mathcal{G}_n(t): 0 \leq t \leq T_{UA}\}$ with its three components, the set-valued process $\{\mathcal{A}_n(t): t \geq 0\}$, the jump process $\{R_m: m \geq 1\}$ and the label process $\{(\beta_m, \gamma_m): m \geq 1\}$, is defined as follows. For each $t \geq 0$, $\mathcal{A}_n(t)$ is a subset of $\{1, 2, \dots\}$ with $\mathcal{A}_n(0) = \{1, 2, \dots, n\}$ and size process $A_n(t) = |\mathcal{A}_n(t)|$. Again, we stop the process at the first time at which $A_n(t) = 1$. We call the particle at that time the *ultimate ancestor* (as in the case when $N < \infty$) and denote the time by T_{UA} ; that is,

$$T_{UA} = \inf\{t \geq 0: A_n(t) = 1\}. \quad (2.4)$$

At times R_m , $m \geq 1$, a coalescing or branching event occurs. Setting $R_0 = 0$, the random times $\{R_m - R_{m-1}\}_{m \geq 1}$ are independent. If $A_n(R_{m-1}) = k$, then $R_m - R_{m-1}$ is exponentially distributed with parameter $\binom{k}{2} + \sigma k/2$; at time R_m a coalescing event occurs with probability $\binom{k}{2} / [\binom{k}{2} + \sigma k/2]$ and a branching event occurs with probability $(\sigma k/2) / [\binom{k}{2} + \sigma k/2]$. If a branching event occurs at time R_m , one particle is chosen at random and its offspring, labelled $n + m$, is added to the collection $\mathcal{A}_n(R_m)$; that is,

$$\mathbb{P}(\beta_m = j, \gamma_m = (0, 0) \mid A_n(R_{m-1}) = k, A_n(R_m) = k + 1) = \frac{1}{k}$$

and

$$\mathcal{A}_n(R_m) = \mathcal{A}_n(R_{m-1}) \cup \{n + m\}.$$

If a coalescing event occurs at time R_m , a pair of particles is chosen at random and the two particles coalesce; that is, if $i < j$,

$$\begin{aligned} \mathbb{P}(\beta_m = 0, \gamma_m = (i, j) \mid A_n(R_{m-1}) = k, A_n(R_m) = k - 1) \\ = \frac{1}{\binom{k}{2}} \end{aligned}$$

and

$$\mathcal{A}_n(R_m) = \mathcal{A}_n(R_{m-1}) \setminus \{j\}.$$

An example of the ancestral selection graph (with one mutation event) is depicted in Figure 3. We call the straight lines *branches*. The three components of the ancestral selection graph are easily recovered from the graph. Conversely, given the three components, one can draw the corresponding graph.

Mutation events occur along the branches of the ancestral selection graph $\mathcal{G}_n(t)$ according to the points of a Poisson process with rate $\theta/2$, independently along each branch of the graph. Mutation events do not occur at points where either a branching or a coalescing event occurs. We denote the mutation process by $\{\mathcal{Y}_n(t): t \geq 0\}$.

We can now formulate the main result of this section.

THEOREM 2.5. *Let $\{\mathcal{G}_n^N(t): 0 \leq t \leq T_{UA}^N\}$ denote the dual process of the biased voter model with mutation process $\{\mathcal{Y}_n^N(t): 0 \leq t \leq T_{UA}^N\}$ on a complete graph of size N , as defined above. Assume $\lambda_1 = N/2$, $\lambda_2 = \lambda_1(1 + s_N)$, $s_N \geq 0$, $u_N \in [0, 1]$, with $Ns_N \rightarrow \sigma \geq 0$ and $Nu_N \rightarrow \theta \geq 0$ as $N \rightarrow \infty$ such that (1.5) holds. Then $\{(\mathcal{G}_n^N(t), \mathcal{Y}_n^N(t)): 0 \leq t \leq T_{UA}^N\}$ converges in distribution as $N \rightarrow \infty$ to $\{(\mathcal{G}_n(t), \mathcal{Y}_n(t)): 0 \leq t \leq T_{UA}\}$.*

The proof of this result is rather technical and is deferred to Appendix A.

COROLLARY 2.6. *The size process $\{A_n(t): 0 \leq t \leq T_{UA}\}$ evolves as follows. If $A_n(t) = k$, $k \geq 2$, then*

$$k \rightarrow \begin{cases} k + 1 & \text{at rate } \frac{\sigma}{2}k \\ k - 1 & \text{at rate } \binom{k}{2} \end{cases}$$

When $\sigma = 0$, the ancestral selection graph reduces to the familiar neutral case where it is called the *n-coalescent*. In the *n-coalescent*, coalescing occurs at rate $\binom{k}{2}$ when the size of the ancestral process is k , and there is no

branching. When $\sigma > 0$, branching occurs at rate $\sigma k/2$ in addition to the coalescing which happens at rate $\binom{k}{2}$, as in the neutral case.

We can summarize our observations as follows: We obtain the ancestral selection graph in the selective case with mutation by first selecting a graph according to the branching/coalescing process. We run this process until the first time its size is equal to 1. We then put mutation events along the branches according to the points of a Poisson process of rate $\theta/2$, independently along each branch of the graph.

It is interesting to note that at first glance, the ancestral selection graph looks like the graph one obtains in the case of recombination without selection (Griffiths 1991, Griffiths and Marjoram 1997); the way the graphs are interpreted, however, is very different in the two cases and there are basically no further similarities between the selection case and the neutral case with recombination.

In the next section we will discuss properties of the ancestral selection graph and describe how to extract the genealogy of a sample from it. Furthermore, we will give an algorithm to simulate the joint distribution of the types of a sample of size n at stationarity.

3. PROPERTIES OF THE ANCESTRAL SELECTION GRAPH

In this section we will analyze the ancestral selection graph defined in the previous section and point out differences with the neutral coalescent.

As noted at the end of the last section, a fundamental difference with the neutral case is that, with selection, branching may occur in addition to coalescing. Since the size of the ancestral process can now exceed the original sample size, it should be clear that branches in the ancestral selection graph no longer necessarily correspond to actual ancestors as in the neutral case. We therefore need to find an interpretation for the ancestral selection graph.

One consequence of the branching is that the most recent common ancestor (MRCA) of a sample need no longer coincide with the first time when the size of the ancestral graph reaches 1, but it will certainly occur by that time. Recall from Section 2 the definition of the time to the ultimate ancestor

$$T_{UA} = \inf\{t: A_n(t) = 1\}. \tag{3.1}$$

Note that the ancestor at time T_{UA} , the ultimate ancestor, is unique. We will first investigate T_{UA} . Let \mathbb{P}_n denote the

law of the ancestral selection graph starting with n ancestors, and let \mathbb{E}_n be the corresponding expectation.

THEOREM 3.2. *For all n , $\mathbb{P}_n(T_{UA} < \infty) = 1$. Moreover, the expected time until the ultimate ancestor starting from a sample of size n , is given by*

$$\mathbb{E}_n T_{UA} = 2 \left(1 - \frac{1}{n}\right) + 2 \sum_{r=1}^{n-1} \frac{1}{r(r+1)} \frac{e^\sigma}{\sigma^{r+1}} \int_0^\sigma t^{r+1} e^{-t} dt. \tag{3.3}$$

In the context of recombination graphs, a quantity similar to T_{UA} can be defined which satisfies (3.3) as well. This was derived in Griffiths (1991) who gave an integral expression for the right hand side of (3.3). Note that $\mathbb{E}_n T_{UA} = 2 - (2/n)$ when $\sigma = 0$, as expected from the corresponding result in the neutral case. (This follows immediately from taking the limit $\sigma \rightarrow 0$ in (3.3) and using L'Hôpital's rule.) Furthermore, $\mathbb{E}_n T_{UA} \rightarrow \infty$ as $\sigma \rightarrow \infty$. We will collect two cases of particular interest in the following corollary, namely the case when the selection coefficient σ is small, and the case when the sample size n is large.

COROLLARY 3.4. *For σ small, we obtain the following power series in σ*

$$\begin{aligned} \mathbb{E}_n T_{UA} = & 2 \left(1 - \frac{1}{n}\right) + \frac{(n-1)(n+2)}{2n(n+1)} \sigma \\ & + \frac{(n-1)(6+4n+n^2)}{9n(n+1)(n+2)} \sigma^2 + O(\sigma^3). \end{aligned} \tag{3.5}$$

For any $\sigma > 0$, we obtain

$$\lim_{n \rightarrow \infty} \mathbb{E}_n T_{UA} = 2 + \frac{2}{\sigma} (-\gamma - \sigma + \text{Ei}(\sigma) - \ln \sigma), \tag{3.6}$$

where $\gamma = 0.577216\dots$ is Euler's constant and $\text{Ei}(\sigma)$ is the second exponential integral function defined by $-\int_{-\sigma}^\infty (e^{-t}/t) dt$ (for $\sigma > 0$), where the principal value of the integral is taken.

Proof of Theorem 3.2. The theorem follows from standard results on birth-death processes (see, e.g., Karlin and Taylor (1975)). Let $\beta_k = k\sigma/2$ and $\alpha_k = \binom{k}{2}$. To show that $\{A_n(t): t \geq 0\}$ hits 1 in finite time with probability one, we need to check the following condition which can be found, for instance, in Theorem 4.7.1 in Karlin and Taylor (1975):

$$\sum_{i=1}^\infty \left(\prod_{k=2}^{i+1} \frac{\alpha_k}{\beta_k} \right) = \infty, \tag{3.7}$$

where empty products have value 1 by convention. It is readily computed that

$$\prod_{k=2}^{i+1} \frac{\alpha_k}{\beta_k} = \frac{i!}{\sigma^i}, \quad i \geq 1, \quad (3.8)$$

from which (3.7) follows immediately.

To derive (3.3), we set $\rho_1 = 1$ and

$$\rho_k = \frac{\beta_2 \cdots \beta_k}{\alpha_2 \cdots \alpha_{k+1}}, \quad k \geq 2.$$

It follows that

$$\rho_k = \frac{2\sigma^{k-1}}{(k+1)!}, \quad k \geq 1. \quad (3.9)$$

The expected waiting time to reach 1 is then given by the following expression which can also be found in Theorem 4.7.1 in Karlin and Taylor (1975):

$$\mathbb{E}_n T_{UA} = \sum_{r=1}^{n-1} \left(\prod_{k=2}^r \frac{\alpha_k}{\beta_k} \right) \sum_{j=r}^{\infty} \rho_j. \quad (3.10)$$

Using (3.8) and (3.9), this can be simplified to

$$\begin{aligned} \mathbb{E}_n T_{UA} &= \sum_{r=1}^{n-1} \frac{(r-1)!}{\sigma^{r-1}} \sum_{j=r}^{\infty} \frac{2\sigma^{j-1}}{(j+1)!} \\ &= 2 \sum_{r=1}^{n-1} \sum_{j \geq 0} \frac{(r-1)!}{(r+j+1)!} \sigma^j. \end{aligned} \quad (3.11)$$

Using MATHEMATICA, we can further simplify the expression on the right-hand side of (3.11) and we obtain

$$\begin{aligned} \mathbb{E}_n T_{UA} &= \sum_{r=1}^{n-1} \frac{2}{r(r+1)} \\ &\quad + 2 \sum_{r=1}^{n-1} \frac{1}{r(r+1)} \frac{e^\sigma}{\sigma^{r+1}} \int_0^\sigma t^{r+1} e^{-t} dt. \end{aligned}$$

Since the first sum

$$\sum_{r=1}^{n-1} \frac{2}{r(r+1)} = 2 \left(1 - \frac{1}{n} \right),$$

(3.3) follows and the theorem is proved. ■

The corollary follows easily from the proof of Theorem 3.2. When σ is small, we can expand $\mathbb{E}_n T_{UA}$ in a power series in σ . Starting at (3.11) and rearranging

TABLE 1

σ	$\lim_{n \rightarrow \infty} \mathbb{E}_n T_{UA}$
0.0	2.0000
0.5	2.2806
1.0	2.6358
5.0	14.199

terms, one arrives after some algebra at the expression given in (3.5). It is straightforward to obtain higher order terms. Our main point is to show how selection affects quantities obtained from the ancestral graph. In this light, terms involving σ in (3.5) can be viewed as ‘‘correction terms.’’ The other case in the corollary is when the sample size n is large. Letting $n \rightarrow \infty$ in (3.11) and using MATHEMATICA, we obtain the expression given in (3.6). Table 1 gives values for $\lim_{n \rightarrow \infty} \mathbb{E}_n T_{UA}$ corresponding to several different values of σ .

The next step is to interpret the ancestral selection graph. As mentioned before, branches in the ancestral selection graph do not necessarily represent lines of descent. Depending on the type of the ultimate ancestor and the mutation events along the branches, certain parts of the ancestral graph may not be accessible to individuals since only individuals of type A_2 may cross 2-arrows. Recall from Section 2 that in the ancestral selection graph ($N \rightarrow \infty$), branching points are formed by incoming 2-arrows and that no mutations occur on any of these 2-arrows.

To pick out the correct genealogical tree for a sample, we will introduce a procedure for simulating the joint distribution of the types of a sample of size n and explain how to obtain the genealogy of the simulated sample.

We defined in Section 1 the process $X(t) = (X_1(t), X_2(t))$ and observed that it reaches a stationary state as $t \rightarrow \infty$; that is, mutation and selection are in equilibrium. If the population has been evolving indefinitely, then the type of an individual sampled at random at any given time has distribution given by the stationary distribution of the process $X(\cdot)$.

To simulate the joint distribution of a sample of size n from a large population ($N \rightarrow \infty$) which is in equilibrium, we proceed as follows:

- (i) Construct the ancestral selection graph according to its distribution given in Section 2, starting with n particles, until the first time its size reaches 1.
- (ii) Choose the type of the ultimate ancestor of the sample according to the stationary distribution of the mutation/selection process.

(iii) Run the mutation process forward along the ancestral selection graph starting at the ultimate ancestor in such a way that the process evolves independently on disjoint branches of the graph and is compatible with the structure of the graph.

This procedure is best explained in an example. See Figure 4 for an illustration when $n = 3$. The first step is to construct the ancestral selection graph. In our example, $A_3(t)$ starts at $A_3(0) = 3$ and goes through the states $3 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1$ with inter-jump times T_3^1, T_2^1, T_3^2 , and T_2^2 . (The subscript k in T_k^j refers to the size of the graph and the superscript j is a counter to distinguish different epochs in which the size of the graph is k .) The choices of which particle branches or which particles coalesce is done uniformly over the particles present just before the time of the jump. Recall that at branching times (such as at time $T_3^1 + T_2^1$ in Figure 4) one of the branches, namely the newly created one, is designated as the incoming branch. In Figure 4, this is the branch that branches off; that is, the branch that connects the nodes labelled $Y_1^{(1)}$ and $Y_2^{(2)}$. We denote by $Y(t)$ the type of a particle which has evolved according to the mutation process for t units of time. This process is used to determine the successive types $Y_l^{(m)}$ along the ancestral selection graph, as described below.

The sample of size 3 is constructed as follows. We determine the type $Y_1^{(0)}$ of the ultimate ancestor according to the stationary distribution of the process $X(\cdot)$. If

the type of the ultimate ancestor is $y \in \{1, 2\}$, then choose $Y_1^{(1)}$ and $Y_2^{(1)}$ independently with distribution that of $Y(T_2^2)$ conditional on $Y_1^{(0)} = y$. Conditional on $Y_1^{(1)} = y_1$ and $Y_2^{(1)} = y_2$, choose $Y_1^{(2)}, Y_2^{(2)}$ and $Y_3^{(2)}$ independently, where $Y_1^{(2)}$ and $Y_2^{(2)}$ have the distribution of $Y(T_3^2)$ conditional on $Y(0) = y_1$, and $Y_3^{(2)}$ has the distribution of $Y(T_3^2)$ conditional on $Y(0) = y_2$. Since the branches associated with $Y_2^{(2)}$ and $Y_3^{(2)}$ coalesce, we need to decide whether $Y_2^{(2)}$ or $Y_3^{(2)}$ will continue. Note that the branch associated with $Y_2^{(2)}$ is the incoming branch. Table 2 tells us what to do. In our example, this means that only if both $Y_2^{(2)}$ and $Y_3^{(2)}$ are of type 1 at the time the two branches come together, will a 1 continue. In all other cases, the continuing type will be a 2. We can then choose $Y_1^{(3)}$ and $Y_2^{(3)}$ according to their respective conditional distributions, and similarly, $Y_1^{(4)}, Y_2^{(4)}$ and $Y_3^{(4)}$. The latter three particles, $Y_1^{(4)}, Y_2^{(4)}$ and $Y_3^{(4)}$, constitute our sample of size 3.

Following the paths of the thus obtained sample backwards to the ultimate ancestor, we obtain the ancestral paths of each individual and hence the *true genealogy* of the sample. This is illustrated in Figures 5a and 5b. Note that the true genealogy depends on the type of the ultimate ancestor. Furthermore, observe that the time to the most recent common ancestor, T_{MRCA} , may occur before T_{UA} and that T_{MRCA} depends on the type of the ultimate ancestor. This dependence on the type of the ultimate ancestor is new and does not occur in the neutral case. In the neutral case, $T_{UA} = T_{MRCA}$ always holds and T_{MRCA} does not depend on the type of the ultimate ancestor (just as T_{UA} does not depend on the type of the ultimate ancestor in the selective case).

We wish to point out that, given an ancestral selection graph with superimposed mutation process, each choice of type for the ultimate ancestor produces a unique ancestral tree which describes the true genealogy and is embedded in the ancestral selection graph. That is, in

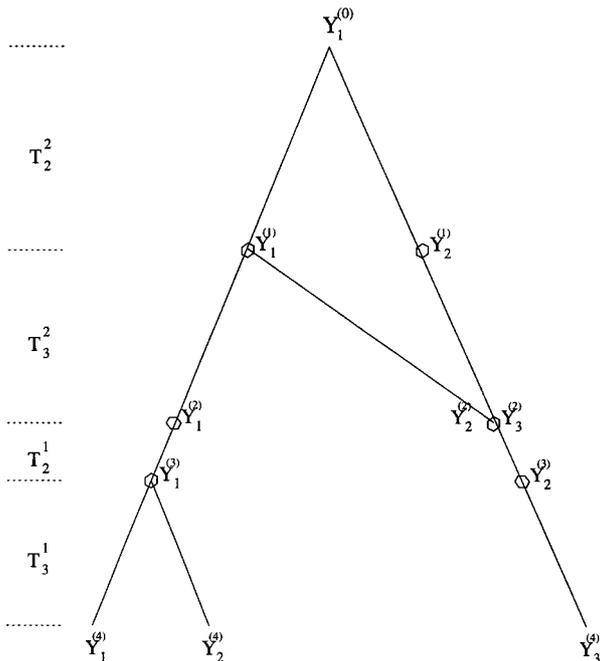


FIG. 4. A realization of the ancestral selection graph when $n = 3$.

TABLE 2

I Is the Incoming Branch

	I	II	III
I	1	1	1
II	1	2	2
III	2	1	2
	2	2	2

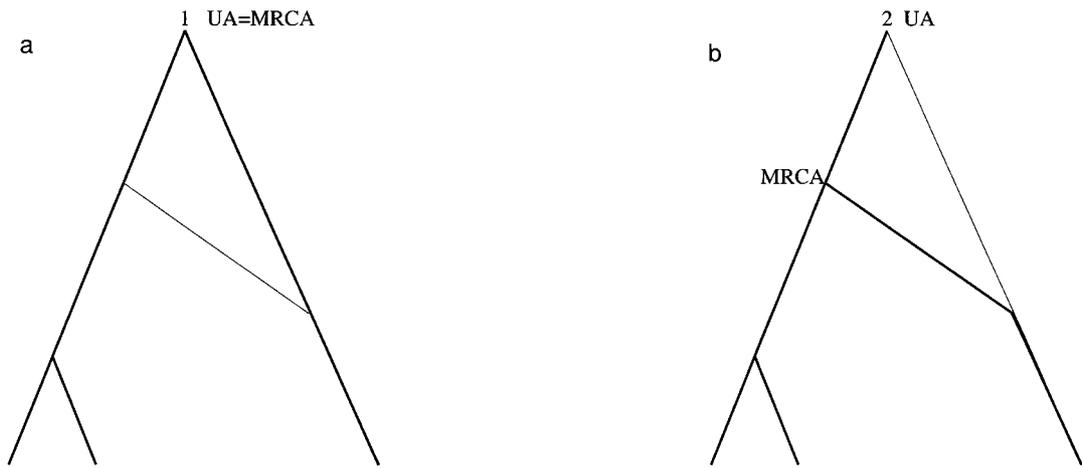


FIG. 5. The ancestral selection graph without mutation events when the ultimate ancestor is (a) of type 1 and (b) of type 2. (Thick lines represent the true genealogy.)

the case of two genotypes, there are exactly two (not necessarily distinct) trees embedded in the ancestral selection graph which describe the genealogy of the constructed sample.

Before we proceed with analyzing T_{MRCA} , we would like to mention how the above procedure needs to be modified in the non-stationary case. In this case, instead of running the mutation/selection process along the branches, simply put mutation events along the branches. Choose the type of the ultimate ancestor according to its distribution and then run the process forward on the graph using Table 2 to decide who will continue when two branches come together forward in time, that is, when we encounter a branching event.

We will first analyze the case when $\theta = 0$, that is, when mutations are absent. In this case, any sample obtained by the above procedure contains only one type, namely, the type of the ultimate ancestor, and it is easy to find the MRCA. If the sample consists only of individuals of type A_1 , the rule at a branching point says that the individual had to come from the continuing branch (since there is also a type A_1 individual on the incoming branch) which implies that in order to find the true genealogy, one simply ignores all branching points and always takes the continuing branches. If the sample consists only of individuals of type A_2 , the rule at a branching point then says that the individual must have come from the incoming branch which implies that one simply follows all incoming branches to obtain the true genealogy.

When determining the MRCA, it is convenient to group particles in the ancestral selection graph in two classes, real particles and virtual particles. We call a particle a *real particle* if it continues along a branch that is part of the true genealogy of the sample, otherwise we

call the particle a *virtual particle*. The rules are now as follows: If a real particle reaches a branching point, it splits into a real particle and into a virtual particle. If a virtual particle reaches a branching point, it splits into two virtual particles. If two particles reach a coalescing point, the resulting particle is real if and only if at least one of the two particles is real, otherwise the resulting particle is virtual. We are now ready to state our next result.

THEOREM 3.12. *Assume $\theta = 0$. Then the distribution of T_{MRCA} does not depend on σ and is, in fact, the same as in the neutral case; that is, $T_{MRCA} = \sum_{k=2}^n T_k$ where T_2, T_3, \dots, T_n are independent and T_k is exponentially distributed with parameter $\binom{k}{2}$.*

The result in Theorem 3.12 is patently obvious from a biological point of view since, if there are no mutations, only one type is present in the population. Since there is no variation in the population, selection cannot have an effect and consequently, the genealogies should not depend on the type. However, this result is not so obvious when looking at the ancestral graph. Given a particular realization of the ancestral graph, the embedded genealogies depend on the type of the UA and are not the same except when there are no branching events in the ancestral selection graph.

Proof of Theorem 3.12. We start with a sample of size n . All n individuals are real particles and of the same type since $\theta = 0$. Each particle branches at rate $\sigma/2$ and each pair of particles coalesces at rate 1. Now, since all particles are of the same type,

$$T_{MRCA} = \inf\{t: \# \text{ of real particles} = 1\}. \quad (3.13)$$

We observe that real particles cannot be created and can only be destroyed if two real particles coalesce. The number of real particles thus goes down by one whenever two real particles coalesce. When $\theta = 0$, all individuals in the sample are of the same type and knowing the types in the sample enables us to designate which particles are real and which ones are virtual.

Let T_k denote the waiting time until the number of real particles drops from k to $k - 1$. Thus, T_2, \dots, T_n are independent and the distribution of T_k is exponential with parameter $\binom{k}{2}$. Since

$$T_{MRCA} = \sum_{k=2}^n T_k,$$

the distribution of T_{MRCA} is independent of σ and is, in fact, the same as in the neutral case. ■

The case $\theta > 0$ is considered in Neuhauser and Krone (1997) where we prove that the conclusion in Theorem 3.12 holds when θ is very large compared to σ . (In fact, the proof for θ very large essentially uses the same ideas as the proof for $\theta = 0$.) Furthermore, computer simulations in that paper indicate that $\mathbb{E}T_{MRCA}$ does not depend very strongly on σ . Both results are for the two-type Moran model considered here. This result is supported by recent simulation results of Golding (1997) which suggest that selection does not have a large effect on the branch lengths of the true genealogy. He obtained this result by simulating the Moran model forward in time, keeping track of all events, and then extracting the genealogy of a sample. It was shown in Kaplan et al. (1988) that, in other models, the branch lengths can change considerably under selection.

One way to get a handle on the distribution of T_{MRCA} is to find bounds on $\mathbb{E}_n T_{MRCA}$. So far we have not been able to obtain useful bounds for $\mathbb{E}_n T_{MRCA}$ in the sense that the bounds would enable us to decide whether Theorem 3.12 is wrong when $\theta > 0$. Of course, there is always the trivial upper bound

$$\mathbb{E}_n T_{MRCA} \leq \mathbb{E}_n T_{UA}. \tag{3.14}$$

Lower bounds can be obtained by selecting ancestral graphs and estimating the expected time until the MRCA. In particular, when σ and n are small, the lower bound is quite good, as expected, since there will be very little branching in the ancestral selection graph.

We denote by $\underline{A} = (a_0, a_1, \dots)$ the path of the jump process $A_n(t)$; that is, a_0, a_1, \dots denote the successive sizes of the ancestral selection graph. Then

$$\begin{aligned} \mathbb{E}_n T_{MRCA} &= \mathbb{E}_n[\mathbb{E}_n(T_{MRCA} \mid \underline{A})] \\ &= \sum_{\underline{a}} \mathbb{E}_n(T_{MRCA} \mid \underline{A} = \underline{a}) \mathbb{P}_n(\underline{A} = \underline{a}) \end{aligned} \tag{3.15}$$

If $\underline{a}_0 = (n, n - 1, n - 2, \dots, 2, 1)$ denotes the “0th term,” i.e., the event of no branching, then

$$\mathbb{P}_n(\underline{A} = \underline{a}_0) = \prod_{k=2}^n \frac{\binom{k}{2}}{\binom{k}{2} + \frac{k\sigma}{2}} \tag{3.16}$$

and

$$\mathbb{E}_n(T_{MRCA} \mid \underline{A} = \underline{a}_0) = \sum_{k=2}^n \frac{1}{\binom{k}{2} + \frac{k\sigma}{2}}. \tag{3.17}$$

Combining (3.16) and (3.17) yields the following lower bound

$$\mathbb{E}_n T_{MRCA} \geq \left(\sum_{k=2}^n \frac{1}{\binom{k}{2} + \frac{k\sigma}{2}} \right) \left(\prod_{k=2}^n \frac{\binom{k}{2}}{\binom{k}{2} + \frac{k\sigma}{2}} \right). \tag{3.18}$$

Since the probability in (3.16) is at most 1 and

$$\mathbb{E}_n(T_{MRCA} \mid \underline{A} = \underline{a}_0) \leq \sum_{k=2}^n \frac{1}{\binom{k}{2}}$$

where the right hand side is the expected time to the MRCA in the neutral case, it follows that the lower bound in (3.18) is at most as large as the expected time to the MRCA in the neutral case and thus does not enable us to decide whether Theorem 3.12 is wrong for $\theta > 0$. Evaluating “higher order terms” in (3.15) allows one to improve the lower bound.

4. IDENTITY BY DESCENT AND LAPLACE TRANSFORM OF THE TIME TO THE MRCA

In this section, we will show how to use the ancestral selection graph to approximate two quantities of interest. These are the probability of identity by descent and the

Laplace transform of the time to the MRCA for two randomly chosen individuals.

We begin by considering the probability that two individuals chosen at random are *identical by descent*, i.e., that there are no mutations in either lineage since their MRCA. We write T_2 for the time to their MRCA. Since the mutation rate $\theta/2$ is the same for either type, mutations occur on each of the branches of the ancestral selection graph according to a Poisson process with rate $\theta/2$.

In the neutral case, the probability that two individuals chosen at random identical by descent is easy to compute. Namely, when $\sigma = 0$,

$$\begin{aligned} \mathbb{P}(\text{identity by descent}) &= \mathbb{E}[\mathbb{P}(\text{identity by descent} \mid T_2)] \\ &= \mathbb{E}[e^{-(\theta/2)2T_2}] \\ &= \int_0^\infty e^{-(\theta/2)2t} e^{-t} dt = \frac{1}{1+\theta}. \end{aligned} \quad (4.1)$$

Computing the probability in (4.1) when $\sigma > 0$ is a lot more difficult since, due to the branching, there is more than one genealogical graph that needs to be considered. In addition, the MRCA depends in general on the type of the UA. We are unable to give a closed expression for this probability when $\sigma > 0$ due to these complications. Instead, we will give a procedure that can be used to obtain a series expansion in σ , when σ is small. Before we explain the procedure, we need to introduce some notation. As in Section 3, we denote by $\underline{a} = (a_0, a_1, \dots)$ the path of the jump process $A_n(t)$. In this section, the sample size n will be equal to 2. We denote by $\mathcal{B}_{\underline{a}}$ the set of all ancestral selection graphs when $n = 2$ with the same path \underline{a} . We suppress the dependence on $n = 2$ in this section and set

$$\mathcal{B} = \bigcup_{\underline{a}} \mathcal{B}_{\underline{a}}. \quad (4.2)$$

We partition each set $\mathcal{B}_{\underline{a}}$ into equivalence classes according to the following equivalence relation: Two graphs from the set $\mathcal{B}_{\underline{a}}$ are in the same equivalence class if their labels (β_m, γ_m) are the same for all $m \geq 1$. That is, all graphs in a given equivalence class share the same topological structure. The times between coalescing or branching events may vary. We denote by $\mathcal{E}_{\underline{a}}$ the set of equivalence classes in $\mathcal{B}_{\underline{a}}$ and call each element in $\mathcal{E}_{\underline{a}}$ a *basic type*. It follows from the construction that all elements of $\mathcal{E}_{\underline{a}}$ are equally likely.

If the random variable G denotes the basic type of the ancestral selection graph, then

$$\begin{aligned} \mathbb{P}(\text{identity by descent}) &= \mathbb{E}[\mathbb{P}(\text{identity by descent} \mid G)] \\ &= \sum_{\underline{a}} \sum_{g \in \mathcal{E}_{\underline{a}}} \mathbb{P}(\text{identity by descent} \mid G = g) \mathbb{P}(G = g). \end{aligned} \quad (4.3)$$

We will use (4.3) to obtain the series expansion in σ mentioned above. It is easy to compute $\mathbb{P}(G = g)$ for any given $g \in \mathcal{E}_{\underline{a}}$. It is not hard, at least in principle, to compute $\mathbb{P}(\text{identity by descent} \mid G = g)$, however, the number of equivalence classes in $\mathcal{E}_{\underline{a}}$ increases quite rapidly as the length of the path \underline{a} increases.

The idea is the following. We order the paths \underline{a} according to their likelihood, starting with the path of largest probability. For σ sufficiently small, the most likely path is $(2, 1)$, the next most likely one is $(2, 3, 2, 1)$, then $(2, 3, 2, 3, 2, 1)$, and so on. It is an easy exercise in combinatorics to compute the number of basic types for any given path \underline{a} . For instance, there is only one basic type that corresponds to the path $(2, 1)$; there are $2\binom{3}{2} = 6$ basic types that correspond to the path $(2, 3, 2, 1)$; there are already $2\binom{3}{2} 2\binom{3}{2} = 36$ basic types that correspond to the path $(2, 3, 2, 3, 2, 1)$; and so on. In Fig. 6 we list the basic types for the paths $(2, 1)$ and $(2, 3, 2, 1)$, where the graphs are interpreted as follows. The left ancestor is ancestor 1, the right ancestor is ancestor 2. If a branching occurs, the parent continues along the straight branch and the offspring, labelled 3, takes the new branch. For instance, in $g(1)$, using the notation introduced in Section 2, the first event at time R_1 is a branching event with $\beta_1 = 1$, $\gamma_1 = (0, 0)$ and the new particle is labelled 3, that is, $\mathcal{A}_2(R_1) = \{1, 2, 3\}$. The next event at time R_2 is a coalescing event with $\beta_2 = 0$, $\gamma_2 = (2, 3)$ and we remove 3 from the set $\mathcal{A}_2(R_1)$, that is, $\mathcal{A}_2(R_2) = \{1, 2\}$. At time R_3 , 1 and 2 coalesce. Note that if the number of branches is larger, we sometimes have to “jump” over branches when drawing basic types. In the cases drawn in Fig. 6, we were able to avoid this by arranging the branches appropriately.

Given the basic type of ancestral selection graph, the mutation events along its branches and the type of the ultimate ancestor, one can determine the MRCA and the embedded genealogical tree. For two individuals to be identical by descent, the branches between the MRCA and the sampled individuals in the embedded genealogical tree need to be free of mutations; any other branches may contain mutations. The mutations in these branches, of course, determine the location of the MRCA and we will have to put restrictions on the number of mutations on these branches.

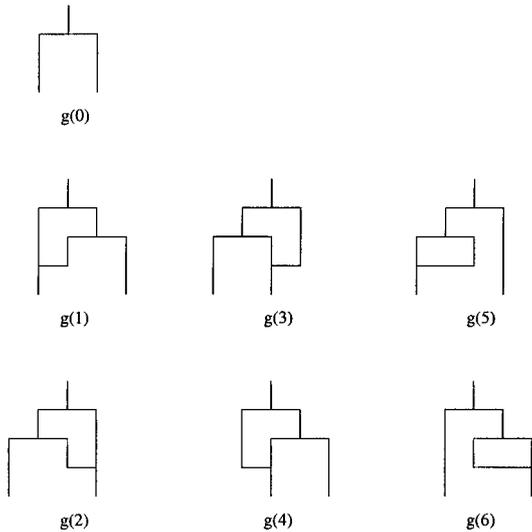


FIG. 6. The basic types corresponding to (2, 1) and (2, 3, 2, 1).

We begin by computing the probabilities in (4.3) for $a = (2, 1)$. In this case, there is only one basic type, therefore, if we denote the basic type corresponding to $a = (2, 1)$ by $g(0)$, then

$$\mathbb{P}(G = g(0)) = \mathbb{P}(A = (2, 1)) = \frac{1}{1 + \sigma} \tag{4.4}$$

since $\mathbb{P}(A = (2, 1))$ is the probability that coalescence occurs before branching; both events occur according to Poisson processes, the former at rate $\binom{2}{2} = 1$, the latter at rate σ . The probability that the two individuals are identical by descent given $G = g(0)$, can be computed as in (4.1), except that T_2 is now exponentially distributed with parameter $1 + \sigma$. Therefore,

$$\begin{aligned} \mathbb{P}(\text{identity by descent} \mid G = g(0)) &= \int_0^\infty e^{-\theta t} (1 + \sigma) e^{-(1 + \sigma)t} dt \\ &= \frac{1 + \sigma}{1 + \sigma + \theta}. \end{aligned} \tag{4.5}$$

Next we compute the probabilities in (4.3) for $a = (2, 3, 2, 1)$. As noted above, there are six basic types which are listed Fig. 6 and labelled $g(1), \dots, g(6)$. As can be seen immediately, the pairs $g(1)$ and $g(2)$, $g(3)$ and $g(4)$, and $g(5)$ and $g(6)$ are mirror images of each other. This will simplify the computation somewhat. Given $A = (2, 3, 2, 1)$, the six basic types are equally likely and, reasoning as in (4.4), we get

$$\begin{aligned} \mathbb{P}(G = g(i)) &= \frac{1}{6} \mathbb{P}(A = (2, 3, 2, 1)) \\ &= \frac{1}{6} \cdot \frac{\sigma}{1 + \sigma} \cdot \frac{3}{3 + \frac{3\sigma}{2}} \cdot \frac{1}{1 + \sigma} \\ &= \frac{1}{6} \cdot \frac{\sigma}{1 + \sigma} \cdot \frac{2}{2 + \sigma} \cdot \frac{1}{1 + \sigma}, \end{aligned} \tag{4.6}$$

for $i = 1, \dots, 6$. To compute the probability that two individuals are identical by descent, we need to consider each pair of symmetric basic types separately starting with the two types of ultimate ancestors. This is best explained by drawing the respective graphs; cf. Fig. 7. For example, the graph $g(1)$ in Fig. 6 can lead to identity by descent in three ways, depending on the type of the ultimate ancestor and the pattern of mutations along the graph. These three sample configurations are shown in $g(1, 1)$, $g(1, 2)$, and $g(1, 3)$ in Fig. 7. Similarly, $g(3, 1)$, $g(3, 2)$, and $g(3, 3)$ are the configurations that yield identity by descent on the graph $g(3)$ of Fig. 6; $g(5, 1)$ and $g(5, 2)$ correspond to $g(5)$ in Fig. 6. Because of the symmetry in the graphs of Fig. 6, we do not write out the sample configurations corresponding to $g(2)$, $g(4)$, and

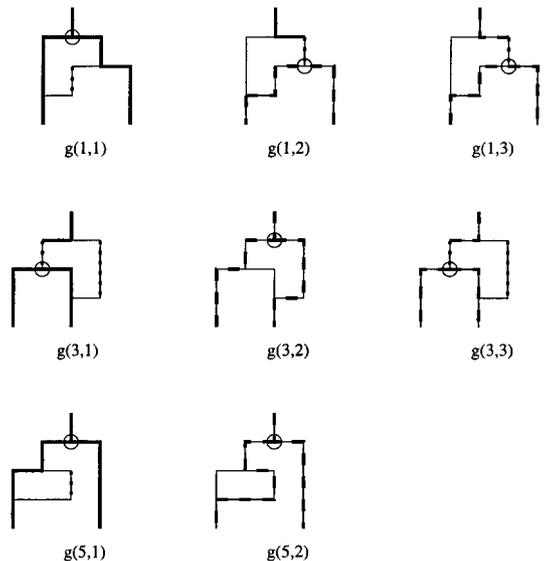


FIG. 7. Embedded genealogical trees for the three basic types. Branches marked with a solid line are used by alleles of type 1 and do not contain mutations. Branches marked with a dashed line are used by alleles of type 2 and do not contain any mutations. Branches marked with a dotted line may contain an even or odd number of mutations depending on the type of the ultimate ancestor. Unmarked branches may contain any number of mutations. The open circle in each graph indicates the location of the MRCA. (See text for further explanation.)

$g(6)$; they will just be the mirror images of the configurations in Fig. 7, and hence have the same probabilities. For ease of drawing, we choose vertical and horizontal segments in the graph. We denote by \tilde{G} the “configuration random variable” corresponding to the graphs $g(i, j)$ in Fig. 7. Then

$\mathbb{P}(\text{identity by descent} \mid G = g(i))$

$$= \sum_{j=1}^{N(i)} \mathbb{P}(\text{identity by descent, } \tilde{G} = g(i, j) \mid G = g(i)),$$

where $g(i, 1), \dots, g(i, N(i))$ are the configurations in Fig. 7 corresponding to the graph $g(i)$.

We explain the procedure on $g(1, 2)$ (which is redisplayed in Fig. 8), since $g(1, 2)$ contains all the features we may encounter.

Starting at the bottom of the graph $g(1, 2)$ in Fig. 8 with two particles (1 and 2), the next event happens at rate $\binom{2}{2} + \sigma = 1 + \sigma$; given that branching occurred, the next event occurs at rate $\binom{3}{2} + 3\sigma/2 = 3 + 3\sigma/2$; given that coalescing occurred, the next event happens at rate $\binom{2}{2} + \sigma = 1 + \sigma$; since coalescing occurs again, we are at the location of the ultimate ancestor. In $g(1, 2)$, the ultimate ancestor is of type 1. If an odd number of mutations occurs along the dotted segment, then the sampled individuals 1 and 2 are identical by descent and of type 2 provided no mutations occur along the dashed lines, regardless of the number of mutations on the unmarked segments. The MRCA is then at the location of the solid dot. (If an even number of mutations had occurred along the dotted segment and no mutations along the dashed segments, and, in addition, the UA had been of type 1, the embedded genealogical tree would have looked like the one in graph $g(1, 1)$ and 1 and 2 would not have been identical by descent.) If we denote the random times

between branching/coalescing events by $T(1)$, $T(2)$, and $T(3)$, as in Fig. 8, and denote by $M(t)$ the number of mutations on a segment of length t , then

$$\begin{aligned} &\mathbb{P}(\text{identity by descent, } \tilde{G} = g(1, 2) \mid G = g(1)) \\ &= \mathbb{E}e^{-(\theta/2) 2(T(1) + T(2))} \mathbb{E}[\mathbb{P}(M(T(3)) \text{ is odd} \mid T(3))] \\ &\quad \times \mathbb{P}(UA = 1) \\ &= \mathbb{E}e^{-\theta T(1)} \mathbb{E}e^{-\theta T(2)} \mathbb{E}[\mathbb{P}(M(T(3)) \text{ is odd} \mid T(3))] \\ &\quad \times \mathbb{P}(UA = 1) \end{aligned} \tag{4.7}$$

We compute each of the first three terms separately.

$$\begin{aligned} \mathbb{E}e^{-\theta T(1)} &= \int_0^\infty e^{-\theta t} (1 + \sigma) e^{-(1 + \sigma)t} dt \\ &= \frac{1 + \sigma}{1 + \sigma + \theta}, \end{aligned} \tag{4.8}$$

$$\begin{aligned} \mathbb{E}e^{-\theta T(2)} &= \int_0^\infty e^{-\theta t} \left(3 + \frac{3\sigma}{2}\right) e^{-(3 + 3\sigma/2)t} dt \\ &= \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} &\mathbb{E}[\mathbb{P}(M(T(3)) \text{ is odd} \mid T(3))] \\ &= (1 + \sigma) \int_0^\infty \mathbb{P}(M(t) \text{ is odd}) e^{-(1 + \sigma)t} dt. \end{aligned} \tag{4.10}$$

Since

$$\begin{aligned} \mathbb{P}(M(t) \text{ is odd}) &= \sum_{k \text{ odd}} \mathbb{P}(M(t) = k) \\ &= \sum_{k \text{ odd}} e^{-(\theta/2)t} \frac{\left(\frac{\theta}{2}t\right)^k}{k!} \\ &= e^{-(\theta/2)t} \frac{e^{(\theta/2)t} - e^{-(\theta/2)t}}{2} \\ &= \frac{1 - e^{-\theta t}}{2}, \end{aligned} \tag{4.11}$$

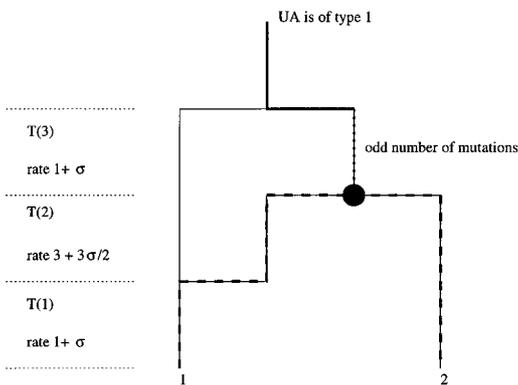


FIG. 8. Graph $g(1, 2)$ redisplayed. See text for explanation.

we get

$$\begin{aligned} & \mathbb{E}[\mathbb{P}(M(T(3)) \text{ is odd} \mid T(3))] \\ &= (1 + \sigma) \int_0^\infty \frac{1 - e^{-\theta t}}{2} e^{-(1+\sigma)t} dt \\ &= \frac{1}{2} (1 + \sigma) \int_0^\infty e^{-(1+\sigma)t} dt \\ &\quad - \frac{1}{2} (1 + \sigma) \int_0^\infty e^{-(1+\sigma+\theta)t} dt \\ &= \frac{1}{2} \frac{1}{2} \cdot \frac{1 + \sigma}{1 + \sigma + \theta} = \frac{\theta}{2(1 + \sigma + \theta)}. \end{aligned} \tag{4.12}$$

Combining (4.8), (4.9) and (4.12), we get

$$\begin{aligned} & \mathbb{P}(\text{identity by descent, } \tilde{G} = g(1, 2) \mid G = g(1)) \\ &= \frac{1 + \sigma}{1 + \sigma + \theta} \cdot \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \\ &\quad \times \frac{\theta}{2(1 + \sigma + \theta)} \cdot \mathbb{P}(UA = 1) \end{aligned} \tag{4.13}$$

The other seven graphs in Fig. 7 are treated in a similar way and the appropriate calculations are relegated to Appendix B. Eq. (4.13) together with the calculations in Appendix B, gives an expression for

$$\sum_{g \in \mathcal{E}_{(2, 3, 2, 1)}} \mathbb{P}(\text{identity by descent} \mid G = g) P(G = g)$$

This expression contains $\mathbb{P}(UA = 1)$. This is the expected value of the relative frequency of A_1 alleles in the diffusive limit under stationarity. In (1.6), the density of the unique stationary distribution was given. We denote the expected value of the relative frequency of A_1 alleles by ρ_1 . It follows from (1.6) that

$$\rho_1 = \mathbb{E}X_1 = \int_0^1 x_1 f(x_1) dx_1 \tag{4.14}$$

with

$$f(x_1) = Kx_1^{\theta-1}(1-x_1)^{\theta-1} e^{-\sigma x_1}, \quad 0 \leq x_1 \leq 1$$

where K is a normalizing constant. The computation of ρ_1 involves Kummer's confluent hypergeometric function

$M(a, b, z)$ which has the following integral representation for $b > a > 0$ (Abramowitz and Stegun (1972)):

$$M(a, b, z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1} dt. \tag{4.15}$$

A direct calculation yields

$$\rho_1 = \frac{1}{2} \frac{M(1 + \theta, 1 + 2\theta, -\sigma)}{M(\theta, 2\theta, -\sigma)}. \tag{4.16}$$

This can be expanded and with the help of MATHEMATICA we get

$$\rho_1 = \frac{1}{2} - \frac{1}{4(1 + 2\theta)} \sigma + O(\sigma^2). \tag{4.17}$$

Combining (4.4) and (4.5) and plugging (4.17) into (B.4) from Appendix B, we can use the MATHEMATICA function Series to find the power series expansion of

$$\sum_{g \in \mathcal{E}_{(2, 1)} \cup \mathcal{E}_{(2, 3, 2, 1)}} \mathbb{P}(\text{identity by descent} \mid G = g) \mathbb{P}(G = g) \tag{4.18}$$

about the point $\sigma = 0$ to the first order. (To compute higher order terms in the series expansion, we would also have to evaluate $\mathbb{P}(\text{identity by descent} \mid G = g)$ for $g \in \mathcal{E}_a$, where a is a path that contains two branching events, since the probability of such a path is of order σ^2 .) Doing the series expansion, we arrive at the following result.

THEOREM 4.19 *For σ small, the probability that two individuals chosen at random are identical by descent is approximately*

$$\frac{1}{1 + \theta} - \frac{\theta(5 + 2\theta)}{4(1 + \theta)^2(3 + \theta)(3 + 2\theta)} \sigma + O(\sigma^2). \tag{4.20}$$

The coefficient of σ is graphed in Fig. 9. Note that as $\sigma \rightarrow 0$, the expression in (4.20) tends to $1/(1 + \theta)$, the probability of identity by descent in the neutral case which we computed in (4.1).

We would like to point out that Theorem 4.19 is the first result in this paper which indicates a difference between the neutral case and the selective case. At least

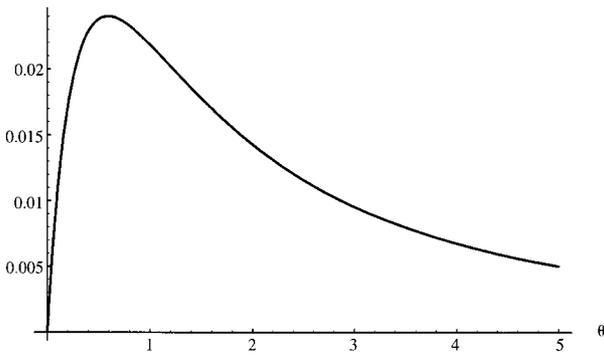


FIG. 9. Graph of the coefficient of σ in (4.25)

for σ small, the probability of identity by descent is a decreasing function in σ . The effect, however, is rather small since the coefficient of σ in (4.24) is quite small as can be seen from Fig. 9.

Another quantity of interest which can be approximated by the above methods is the Laplace transform of T_2 . In the neutral case, T_2 has an exponential distribution with mean 1, so

$$\mathbb{E}(e^{-\alpha T_2}) = \frac{1}{1 + \alpha}.$$

Again, selection complicates things because there are different ancestral selection graphs which, together with the distribution of mutations and the type of the ultimate ancestor, can give different distributions for T_2 . We begin with

$$\mathbb{E}(e^{-\alpha T_2}) = \sum_g \mathbb{E}[e^{-\alpha T_2} | G = g] \mathbb{P}(G = g), \quad (4.21)$$

where the sum is over all possible ancestral selection graphs starting with a sample of size 2 as in (4.3). As above, for small σ , we can order the graphs according to their likelihoods. We will again use only the graphs $g(0)$, $g(1)$, ..., $g(6)$ in Fig. 6 to carry out our approximation.

For $g(0)$, there are no branching events and T_2 is exponential with mean $(1 + \sigma)^{-1}$. So, no matter what the type of the UA or the distribution of mutations along the branches,

$$\begin{aligned} \mathbb{E}[e^{-\alpha T_2} | G = g(0)] &= \int_0^{\infty} e^{-\alpha t} (1 + \sigma) e^{-(1 + \sigma)t} dt \\ &= \frac{1 + \sigma}{1 + \sigma + \alpha}, \end{aligned} \quad (4.22)$$

as in (4.5).

We now consider the graphs $g(1)$, ..., $g(6)$ in Fig. 6 which involve a single branching event. For each of these graphs, the number of ancestors has the pattern $a = (2, 3, 2, 1)$. As in Fig. 8, let us use $T(1)$, $T(2)$, and $T(3)$ to denote the times between jumps. Thus $T(1)$ and $T(3)$ are exponential random variables with mean $(1 + \sigma)^{-1}$, and $T(2)$ is exponential with mean $(3 + 3\sigma/2)^{-1}$. The value of $\mathbb{E}[e^{-\alpha T_2} | G = g(i)]$ will be determined by the location of the MRCA in the embedded tree. This, in turn, is completely determined by the type of the UA and the number of mutations on the path from the UA to the branching point going through the incoming branch. All other mutations are irrelevant for this calculation. Once again, due to symmetry, we need only consider $g(1)$, $g(3)$, and $g(5)$.

First consider $g(1)$. There are two possible locations for the MRCA: at the UA, in which case $T_2 = T(1) + T(2) + T(3)$, or at the location of the first coalescence, in which case $T_2 = T(1) + T(2)$. The first case will occur if the incoming branch is not in the true tree; the second case will occur if the incoming branch is used in the true tree.

Thus we have $T_2 = T(1) + T(2) + T(3)$ if

- (a) UA = 1, even number of mutations for time $T(3) + T(2)$; or
- (b) UA = 2, odd number of mutations for time $T(3) + T(2)$.

We will have $T_2 = T(1) + T(2)$ if

- (a') UA = 1, odd number of mutations for time $T(3) + T(2)$; or
- (b') UA = 2, even number of mutations for time $T(3) + T(2)$.

To compute the probabilities of these events, recall that $M(t)$ denotes the number of mutations on a path of length t , and that

$$\mathbb{P}(M(t) \text{ is odd}) = \frac{1 - e^{-\theta t}}{2}$$

and

$$\mathbb{P}(M(t) \text{ is even}) = \frac{1 + e^{-\theta t}}{2}.$$

We have

$$\begin{aligned} &\mathbb{P}((a) \text{ or } (b) | G = g(1)) \\ &= \mathbb{P}(UA = 1) \mathbb{P}(M(T(3) + T(2)) \text{ is even}) \\ &\quad + \mathbb{P}(UA = 2) \mathbb{P}(M(T(3) + T(2)) \text{ is odd}). \end{aligned} \quad (4.23)$$

Similarly,

$$\begin{aligned} & \mathbb{P}(a') \text{ or } (b') \mid G = g(1)) \\ &= \mathbb{P}(UA = 1) \mathbb{P}(M(T(3) + T(2)) \text{ is odd}) \\ &+ \mathbb{P}(UA = 2) \mathbb{P}(M(T(3) + T(2)) \text{ is even}). \end{aligned} \quad (4.24)$$

We perform the relevant calculations in Appendix C.

Combining (C.3), (C.4), and (C.5) from Appendix C with (4.4), (4.6), and (4.18) and simplifying, we can again use the MATHEMATICA function `Series` to find the power series expansion of

$$\begin{aligned} & \sum_{i=0}^6 \mathbb{E}[e^{-\alpha T_2} \mid G = g(i)] \mathbb{P}(G = g(i)) \\ &= \frac{1}{1 + \sigma + \alpha} + \frac{1}{3} \left(\frac{\sigma}{1 + \sigma} \right) \left(\frac{2}{2 + \sigma} \right) \left(\frac{1}{1 + \sigma} \right) \\ &\quad \times \left(\frac{1 + \sigma}{1 + \sigma + \alpha} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \alpha} \right) \\ &\quad \times \left(2 \frac{1 + \sigma}{1 + \sigma + \alpha} + 1 \right) \end{aligned} \quad (4.25)$$

about the point $\sigma = 0$ to the first order. We arrive at the following result.

THEOREM 4.26. *For small σ , the Laplace transform of T_2 , the time of the MRCA of a sample of size 2, is approximately*

$$\frac{1}{1 + \alpha} + O(\sigma^2). \quad (4.27)$$

Note that the linear term in σ vanishes in the expansion of the Laplace transform of T_2 . This is an indication that the time to the MRCA does not depend on σ very strongly when σ is small. The result of this series expansion is somewhat reminiscent of the series expansion for the expected homozygosity (Watterson, 1978) where the linear term in σ vanishes as well for the model considered here.

5. RECURSIONS FOR SAMPLING DISTRIBUTIONS

The algorithm for simulating samples, described in Section 3, is not very useful if one wishes to estimate the probability of a particular sample configuration since the

state space is too large. Griffiths and Tavaré (1994) introduced an efficient computational algorithm for computing the probability distribution for samples taken from infinite populations that evolve according to certain neutral population models. The algorithm is based on a recursion satisfied by this sampling distribution, which allows one, at least in principle, to compute sampling probabilities exactly. In practice, this recursion is usually too large to solve directly and one needs to resort to computer simulations. Griffiths and Tavaré (1994) developed a scheme using Markov chain Monte Carlo methods for approximating these sampling probabilities via similar recursions for the neutral case. We will extend their recursion to the selective case.

We say that a sample of n genes taken from a population with allele frequencies (X_1, X_2) has *type configuration* $\mathbf{n} = (n_1, n_2)$, if the sample consists of n_1 genes of type A_1 and n_2 genes of type A_2 with $n = n_1 + n_2$. We first recall the density of the unique stationary distribution which is known as Wright's formula (Wright 1949; Watterson 1977; see also Lemma 4.1 in Ethier and Kurtz 1994) and which was given in (1.6), namely,

$$h(x_1, x_2) = Kx_1^{\theta-1}x_2^{\theta-1}e^{\sigma x_2}.$$

If (X_1, X_2) is distributed according to the stationary distribution and we denote by $q(\mathbf{n})$ the probability that a sample of n genes taken from a population in equilibrium has type configuration $\mathbf{n} = (n_1, n_2)$, then it follows that

$$q(\mathbf{n}) = \frac{n!}{n_1!n_2!} \mathbb{E}_n(X_1^{n_1}X_2^{n_2}). \quad (5.1)$$

The probability $q(\mathbf{n})$ satisfies a fundamental recursion. This is the content of the next theorem.

THEOREM 5.2. *The probability $q(\mathbf{n})$ satisfies*

$$\begin{aligned} q(\mathbf{n}) &= \frac{\theta}{n + \theta + \sigma - 1} \left(\sum_{\substack{i, j \in \{1, 2\} \\ n_j > 0, i \neq j}} \frac{n_i + 1}{n} q(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) \right) \\ &+ \frac{n - 1}{n + \theta + \sigma - 1} \left(\sum_{\substack{j \in \{1, 2\} \\ n_j > 0}} \frac{n_j - 1}{n - 1} q(\mathbf{n} - \mathbf{e}_j) \right) \\ &+ \frac{\sigma}{n + \theta + \sigma - 1} \left(\sum_{j \in \{1, 2\}} \frac{n_j(n_j + 1)}{n(n + 1)} q(\mathbf{n} + \mathbf{e}_j) \right) \\ &+ 2 \frac{n_2(n_1 + 1)}{n(n + 1)} q(\mathbf{n} + \mathbf{e}_1) \end{aligned} \quad (5.3)$$

where \mathbf{e}_1 and \mathbf{e}_2 are the unit vectors. The boundary conditions to solve (5.3) have the form

$$q(\mathbf{e}_i) = \rho_i, \quad i = 1, 2, \quad (5.4)$$

where ρ_i is the probability that the ultimate ancestor is of type A_i .

The recursion in (5.3) can then be used as in Griffiths and Tavaré (1994) to obtain approximations of sampling probabilities using Markov chain Monte Carlo methods. Of course, the probability of a sample type configuration can be calculated (numerically) directly if the stationary distribution is known in closed form as in (1.6). However, if the scheme is more complex and the analogue of (1.6) is not known (as in the case when the mutant type depends on the parent type), then the Griffiths–Tavaré technique is useful. We will give two proofs of Theorem 5.2. The first proof is purely analytical, the second proof uses the structure of the ancestral selection graph.

1st Proof of Theorem 5.2. The first step is to define the generator of the diffusion process one obtains in the limit $N \rightarrow \infty$ with the scaling given in (1.3)–(1.5). The generator for the diffusion process can be found, for instance, in Ethier and Kurtz (1994). The state space of the diffusion process is $\mathcal{A}_2 = \{(x_1, x_2) \in [0, 1]^2 : x_1 + x_2 = 1\}$. The generator in Ethier and Kurtz (1994) is for a diploid model; however, the haploid model is equivalent to the diploid model with genic selection. Setting $\gamma_1 = \gamma_2 = \theta$ and $\sigma_{kl} = \sigma_k + \sigma_l$ with $\sigma_1 = 0$ and $\sigma_2 = \sigma/2$ in Eqs. (4.10) and (4.11) of Ethier and Kurtz (1994), we can write the generator of the diffusion process as

$$\begin{aligned} L = & \frac{1}{2} x_1(1-x_1) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} x_2(1-x_2) \frac{\partial^2}{\partial x_2^2} - x_1 x_2 \frac{\partial^2}{\partial x_1 \partial x_2} \\ & + \frac{\theta}{2} (x_2 - x_1) \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} \right) \\ & + \frac{\sigma}{2} x_1 x_2 \left(\frac{\partial}{\partial x_2} - \frac{\partial}{\partial x_1} \right). \end{aligned} \quad (5.5)$$

If (X_1, X_2) is distributed according to the stationary distribution, then

$$\mathbb{E}_n(LX_1^n X_2^n) = 0. \quad (5.6)$$

Evaluating (5.6) and using the definition of $q(\mathbf{n})$ in (5.6) leads to the recursion in (5.3). Indeed, after some algebra, one obtains

$$\begin{aligned} q(\mathbf{n}) & \left[\frac{1}{2} n(n-1) + \frac{\theta}{2} n + \frac{\sigma}{2} n \right] \\ & = \frac{1}{2} n(n_1-1) q(\mathbf{n} - \mathbf{e}_1) + \frac{1}{2} n(n_2-1) q(\mathbf{n} - \mathbf{e}_2) \\ & \quad + \frac{\theta}{2} [(n_2+1) q(\mathbf{n} + \mathbf{e}_2 - \mathbf{e}_1) \\ & \quad + (n_1+1) q(\mathbf{n} + \mathbf{e}_1 - \mathbf{e}_2)] \\ & \quad + \frac{\sigma}{2(n+1)} [n_1(n_1+1) q(\mathbf{n} + \mathbf{e}_1) \\ & \quad + 2n_2(n_1+1) q(\mathbf{n} + \mathbf{e}_1) \\ & \quad + n_2(n_2+1) q(\mathbf{n} + \mathbf{e}_2)]. \end{aligned} \quad (5.7)$$

The only tricky part is the selection term. Applying that term to $x_1^n x_2^n$ yields

$$\begin{aligned} & \frac{\sigma}{2} x_1 x_2 \left(\frac{\partial}{\partial x_2} - \frac{\partial}{\partial x_1} \right) x_1^n x_2^n \\ & = \frac{\sigma}{2} (n_2 x_1^{n_1+1} x_2^{n_2} - n_1 x_1^n x_2^{n_2+1}). \end{aligned} \quad (5.8)$$

We will now manipulate the right hand side of (5.8). First, we use $x_2 = 1 - x_1$ and add/subtract the term $n_2 x_1^{n_1+1} x_2^n$. This yields

$$\begin{aligned} & n_2 x_1^{n_1+1} x_2^n - n_1 x_1^n x_2^{n_2+1} \\ & = n_2 x_1^{n_1+1} x_2^n - n_1 x_1^n (1-x_1) x_2^n \\ & \quad - n_2 x_1^{n_1+1} x_2^n + n_2 x_1^{n_1+1} x_2^n \\ & = n_2 x_1^{n_1+1} x_2^n - n_1 x_1^n (1-x_1) x_2^n \\ & \quad - n_2 x_1^n (1-x_2) x_2^n + n_2 x_1^{n_1+1} x_2^n \end{aligned}$$

where we used $x_1 = 1 - x_2$ in the last step. Collecting positive and negative terms separately, this becomes

$$\begin{aligned} & = n_2 x_1^{n_1+1} x_2^n + n_1 x_1^{n_1+1} x_2^n + n_2 x_1^n x_2^{n_2+1} + n_2 x_1^{n_1+1} x_2^n \\ & \quad - n_1 x_1^n x_2^n - n_2 x_1^n x_2^n. \end{aligned} \quad (5.9)$$

We still need to multiply (5.9) by $(\sigma/2)(n!/(n! n_2!))$ and then simplify terms before we get the selection term in (5.7). We demonstrate this on the first term in (5.9):

$$\begin{aligned} \frac{\sigma}{2} \frac{n!}{n_1! n_2!} n_2 x_1^{n_1+1} x_2^{n_2} &= \frac{\sigma}{2} \frac{n!}{n_1! n_2} n_2 \frac{n+1}{n+1} \frac{n_1+1}{n_1+1} x_1^{n_1+1} x_2^{n_2} \\ &= \frac{\sigma}{2} \frac{n_1+1}{n+1} n_2 \frac{(n+1)!}{(n_1+1)! n_2!} x_1^{n_1+1} x_2^{n_2} \\ &= \frac{\sigma}{2} \frac{n_1+1}{n+1} n_2 q(\mathbf{n} + \mathbf{e}_1), \end{aligned}$$

which is part of the middle term in the selection term in (5.7). The next three terms are treated similarly and we omit the details. The last two terms in (5.9) can be combined to $-n x_1^{n_1} x_2^{n_2}$. After multiplying this by $(\sigma/2)(n!/(n_1! n_2!))$, one gets $(-\sigma/2) n q(\mathbf{n})$ which is moved to the left side of (5.7). A minor algebraic manipulation then yields (5.3). ■

The second proof of Theorem 5.2 uses the structure of the ancestral selection graph. This approach has been used a number of times in the neutral case, see for instance, Sawyer, Dykhuizen and Hartl (1987), Lundstrom (1990), and Griffiths and Tavaré (1994). As we will see, this works as well when selection is present.

2nd Proof of Theorem 5.2. To derive (5.3), consider the first event back in time that happens in the ancestral selection graph. Relative rates of mutation, coalescence, and selection (branching) for n alleles are $n\theta/2 : n(n-1)/2 : n\sigma/2$.

The probability that the first event is a mutation is $\theta/(n+\theta+\sigma-1)$. To obtain the configuration \mathbf{n} after mutation, the configuration before mutation must have been $\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j$, $i \neq j$, and a transition $i \rightarrow j$ must have taken place. Note that, before the mutation, there were $n_i + 1$ alleles of type A_i and one of them is chosen at random from among the n alleles present in the graph. This gives the factor $(n_i + 1)/n$. Note that the assumption of stationarity justifies the use of $q(\cdot)$ before and after the mutation event.

The probability that the first event is a coalescence is $(n-1)/(n+\theta+\sigma-1)$. To obtain the configuration \mathbf{n} , the configuration before the coalescence event must have been $\mathbf{n} - \mathbf{e}_j$, and the ancestral line involved must have been of type A_j . Note that there were $n_j - 1$ alleles of type A_j among the $n-1$ alleles present in the graph and one of them was chosen at random to reproduce, which gives the factor $(n_j - 1)/(n-1)$.

The probability that the first event back is a branching event, is $\sigma/(n+\theta+\sigma-1)$. To obtain the configuration \mathbf{n} , we need to look at Table 2 which tells us that the configuration must have been either in state $\mathbf{n} + \mathbf{e}_j$, $j = 1, 2$, and the two ancestral lines involved were of the same

type A_j (that is, one of the two particles is lost in the branching event and the configuration changes from $\mathbf{n} + \mathbf{e}_j$ to \mathbf{n}), or in state $\mathbf{n} + \mathbf{e}_1$ and the ancestral lines involved were of different types (that is, the A_1 allele is lost since the A_2 allele wins according to Table 2 and the configuration changes from $\mathbf{n} + \mathbf{e}_1$ to \mathbf{n}). In the first case, we pick two alleles of the same type out of $n+1$ alleles; this has probability $\binom{n_j+1}{2} / \binom{n+1}{2} = (n_j(n_j+1)) / (n(n+1))$. In the second case, we pick two different alleles and, since only type A_2 survives, we pick one of the n_1+1 A_1 alleles and one of the n_2 A_2 alleles; this has probability $(n_2(n_1+1)) / \binom{n+1}{2} = (2n_2(n_1+1)) / (n(n+1))$. ■

6. OTHER MODELS IN POPULATION GENETICS

In the previous sections, we focussed entirely on the Moran model with symmetric mutation and so-called fertility selection as defined in (1.1). In this section, we will discuss other models and show that they yield essentially the same ancestral selection graph.

The first obvious generation is to allow for K alleles with respective reproduction parameters $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_K$ and where the probability that the offspring of an individual of type A_i is a mutant, is denoted by $p_N(i)$, $i = 1, 2, \dots, K$ and given that the offspring is a mutant, a mutation from type i to type j , $1 \leq j \leq K$, occurs according to a transition probability $\Gamma_i = (\gamma_{ij})_{1 \leq j \leq K}$. This defines a transition matrix $\Gamma = (\gamma_{ij})_{0 \leq i, j \leq K}$. That is, an individual of type A_i gives birth to a new individual at rate λ_i and the type of the offspring will be A_j with probability $p_N(i) \gamma_{ij}$.

This generalization is easily incorporated into the graphical representation and the structure of the ancestral selection process follows in a straightforward way: We put down arrows at rate λ_K and then use a thinning process similar to the one in Section 2 to obtain a hierarchy of arrows starting at 1, that are successively open to the different types; that is, 1-arrows appear at rate λ_1 and are open individuals of all types, 2-arrows appear at rate $\lambda_2 - \lambda_1$ and are open to individuals of type 2 or higher, 3-arrows appear at rate $\lambda_3 - \lambda_2$ and are open to individuals of type 3 or higher, and so on. If we set

$$\lambda_i = \lambda_1(1 + s_N(i)), \quad i = 2, 3, \dots, K \quad (6.1)$$

with $\lambda_1 = N/2$ and $N s_N(i) \rightarrow \sigma_i$ as $N \rightarrow \infty$, we obtain as the limiting object a coalescing, branching graph in which

incoming branches are labelled from 1 to K according to the corresponding arrows in the graphical representation. A look-up table similar to Table 2 can then be constructed. We can then superimpose the mutation process as in the neutral case since different branches in the ancestral selection graph evolve independently of each other. That is, for $i = 1, 2, \dots, K$, we assume

$$|Np_N(i) - \theta_i| \rightarrow 0, \quad \text{as } N \rightarrow \infty, \quad (6.2)$$

where the rate of convergence in (6.2) again needs to be faster than some negative power of N . In the limiting object, mutations occur at rate $\theta_i/2$ along the branches, and the types are then chosen according to the transition matrix Γ defined above.

Having defined the ancestral selection graph for a K allele model, we can extend this immediately to *infinitely-many-alleles* models with d selection classes. This class of models was introduced in Ethier and Kurtz (1987, 1994); see also Joyce and Tavaré (1995) for an application. In these models, each gene is assigned a type $x \in [0, 1)$. Alleles are classified into one of $d \geq 2$ classes and the selection coefficients are constant for genes in a given class. To do this, we define $c_0 = 0$, $c_d = 1$ and numbers $c_1 < c_2 < \dots < c_{d-1} \in (0, 1)$ and set

$$I_j = [c_{j-1}, c_j), \quad j = 1, 2, \dots, d. \quad (6.3)$$

We say that an allele of type $x \in [0, 1)$ falls into the j th selection class if $x \in I_j$. The reproduction rate of an individual in the j th selection class is then defined as

$$\lambda_j = \lambda_1(1 + s_N(j)), \quad j = 1, 2, \dots, d \quad (6.4)$$

with $0 \leq s_N(1) \leq s_N(2) \leq \dots \leq s_N(d)$. The type of the offspring is then chosen according to the mutation process; that is, the offspring is either the same type as the parent if no mutation occurs, or of a type that has not been previously observed in the population. This can be done by choosing a number $x \in [0, 1)$ according to some continuous distribution. The value x will denote the new mutant type (which will be different from all previous types with probability 1) and the selection class is determined by the interval into which x falls. One can then proceed as in the K allele case. The ancestral selection graph again has the same coalescing, branching structure.

So far, we have only discussed fertility selection, that is, the reproduction success depends on the type of the parent. One can also consider *viability selection*, in which the probability that an offspring can replace an individual depends on the type of the individual chosen

to be replaced. To define viability selection, we will again restrict ourselves to the two-allele case. We change the dynamics in the following way:

An individual of type A_i , $i = 1, 2$, dies at rate δ_i with $\delta_1 \geq \delta_2$. Immediately before death, an individual is chosen at random from the population which will replace the dying individual.

This can be implemented into the graphical representation in the following way. There are again two types of arrows: arrows that are labelled with a δ at the tip of the arrow and which appear at rate δ_2 , and unlabelled arrows which appear at rate $\delta_1 - \delta_2$. If an individual encounters the tip of an arrow labelled with a δ , it dies and is replaced by the individual from which the arrow emanated. If an individual encounters the tip of an unlabelled arrow, it is only replaced if it is of type 1 and it ignores the arrow if it is of type 2. It is easy to see that after rescaling the parameters appropriately and letting $N \rightarrow \infty$, the ancestral process yields the same coalescing, branching graph as before. In addition, the look-up table defined in Table 2, is the same.

As noted in Moran (1962), the generator for the associated diffusion is the same in the case of viability selection and fertility selection. This implies that the recursion, obtained in Section 5 for fertility selection, also holds for viability selection.

We would also like to point out that the generators of many exchangeable models in population genetics converge to the same generator in the diffusive limit. (Exchangeable models were introduced by Cannings (1974). In short, exchangeability means that the individuals of a generation need not be labelled in a particular way.) The most celebrated model which yields the same diffusion generator (5.5) after appropriate scaling, is the *Wright-Fisher model* (Fisher 1930, Wright 1931) in which time is discrete; that is, genes in generation $t + 1$ are derived by sampling with replacement from the genes of generation t . (The Moran model is a continuous-time version of the Wright-Fisher model.) In Theorem 5.2, we obtain a recursion for the quantity $q(\mathbf{n})$ defined in Section 5. We showed two proofs of this theorem. The second proof was based on knowing the ancestral selection graph, whereas the first proof merely used the generator for the diffusion. Reversing the argument used in the second proof, we can also read off the ancestral process from the recursion. (This is a well-known technique in interacting particle systems to compute coalescing or annihilating dual processes if they exist (see, e.g., Liggett 1985).) Since many exchangeable processes possess the same diffusion limit, this shows that the ancestral selection graph is also robust under change of models. The

phenomenon of robustness was already observed by Kingman (1982b). In Neuhauser and Krone (1997), we discuss the ancestral selection graph for other models as well. These include infinite sites, DNA sequence models, and diploid models.

As mentioned in Section 1, Kaplan, Darden, and Hudson (1988) investigated the genealogical history of a sample of genes taken from a population that evolved according to a selective model. (In Hudson and Kaplan (1988), this was extended to include recombination.) They essentially used Kingman's coalescent process and adjusted for selection by assuming that the distributions of the frequencies of alleles were known throughout the ancestral generations. They thus arrived at the conclusion that "For models with selection, in contrast to models without selection, the distribution of this process, the coalescent process, depends on the distribution of the frequencies of alleles in the ancestral generations." In contrast, the ancestral selection graph we obtained does not require this knowledge. All one needs to know is the distribution of the frequencies of alleles at the time of the ultimate ancestor. It is not surprising that one needs to know the distribution of the frequencies of alleles throughout the ancestral history if one simply uses Kingman's coalescent to study the ancestral history of genes in the selective case, since Kingman's coalescent holds strictly only in the neutral case. The additional branches that are contained in the ancestral selection graph, when $\sigma > 0$, favor the advantageous allele. Using the correct ancestral selection graph allows one to find the embedded genealogy and does not require knowledge of the frequency distribution at all times.

APPENDIX A: PROOF OF THEOREM 2.5

We demonstrate Theorem 2.5, that is, the convergence in distribution of $\mathcal{G}_n^N(t)$ to $\mathcal{G}_n(t)$, as $N \rightarrow \infty$, by coupling the two processes and showing that the probability that the two processes are identical goes to 1 as $N \rightarrow \infty$. We start the processes with $\mathcal{A}_n^{N,1}(0) = \mathcal{A}_n(0) = \{1, 2, \dots, n\}$ and $\mathcal{A}_n^{N,2}(0) = \emptyset$ and couple the times R_m^N and R_m , $m \geq 1$, and the labels (β_m^N, γ_m^N) and (β_m, γ_m) , $m \geq 1$, so that the respective processes agree as much as possible.

To obtain bounds on the size and the time to the ultimate ancestor in both processes, we introduce the following continuous time Markov chain on $\{1, 2, \dots\}$, denoted by $Q_n(t)$, with $Q_n(0) = n$, which dominates both $A_n^N(t)$ and $A_n(t)$ for sufficiently large N , in the sense

$$\mathbb{P}(Q_n(t) \geq A_n^N(t) \vee A_n(t)) = 1, \quad \text{for } N \geq N_0, \quad (\text{A.1})$$

where $Q_n(0) = A_n^N(0) = A_n(0) = n$ and N_0 only depends on the convergence rate in (1.5). If $Q_n(t) = k$, then the transitions of $Q_n(t)$ are as follows:

$$k \rightarrow \begin{cases} k+1 & \text{at rate } \sigma k \\ k-1 & \text{at rate } \binom{k}{2}. \end{cases} \quad (\text{A.2})$$

Property (A.1) follows immediately from (A.2) by the usual coupling since $Ns_N/2 \leq \sigma$ for N sufficiently large, and $\sigma/2 \leq \sigma$ trivially holds. We set

$$\tau = \inf\{t : Q_n(t) = 1\} \quad (\text{A.3})$$

and

$$M_n = \max\{Q_n(t) : 0 \leq t \leq \tau\}$$

and stop $Q_n(t)$ at time τ . It is standard fare to show that there exists a constant $K < \infty$, depending on n but independent of t and N , so that

$$\mathbb{E}\tau \leq K. \quad (\text{A.4})$$

The distribution of M_n can be computed from a recursion for $\mathbb{P}(M_n \leq k)$. This was done in Griffiths (1991) and one obtains

$$\mathbb{P}(M_n \leq k) = \frac{\sum_{j=n-1}^{k-1} j! (2\sigma)^{-j}}{\sum_{j=0}^{k-1} j! (2\sigma)^{-j}}.$$

This immediately yields that there exists a constant $K < \infty$, depending on n but independent of t and N , so that

$$\mathbb{E}M_n \leq K. \quad (\text{A.5})$$

We can choose K to be the same in (A.4) and (A.5). It is immediate from (A.1) that

$$\mathbb{P}(\tau \geq T_{UA}^N \vee T_{UA}) = 1 \quad (\text{A.6})$$

for N sufficiently large. We also need to introduce the maximum size of $\mathcal{G}_n^N(t)$ and $\mathcal{G}_n(t)$. We set

$$W_n^N = \max\{A_n^N(t) : 0 \leq t \leq T_{UA}^N\}$$

and

$$W_n = \max\{A_n(t) : 0 \leq t \leq T_{UA}\}.$$

We can now define the bad events which cause $\mathcal{G}_n^N(t)$ and $\mathcal{G}_n(t)$ to differ. Fix some $\alpha < (\gamma/2) \wedge (1/3)$ and define

$$F_1 = \{T_{UA}^N \vee T_{UA} > N^\alpha\}$$

$$F_2 = \{W_n^N \vee W_n > N^\alpha\} \cap F_1^c$$

$$F_3 = \{\text{a collision occurs in } \mathcal{G}_n^N(t) \text{ at some } t \leq N^\alpha\} \\ \cap F_1^c \cap F_2^c$$

$$F_4 = \{\text{a branching event occurs in exactly one of the} \\ \text{two processes } \mathcal{G}_n^N(t) \text{ or } \mathcal{G}_n(t) \text{ at some } t \leq N^\alpha\} \\ \cap F_1^c \cap F_2^c \cap F_3^c.$$

To estimate $\mathbb{P}(F_1)$, we use Markov's inequality together with (A.4) and (A.6) to obtain

$$\mathbb{P}(F_1) \leq \mathbb{P}(\tau > N^\alpha) \leq N^{-\alpha} \mathbb{E}\tau \leq KN^{-\alpha} \quad (\text{A.7})$$

for N sufficiently large. To estimate $\mathbb{P}(F_2)$, we use Markov's inequality together with (A.5). We obtain

$$\mathbb{P}(M_n > N^\alpha) \leq N^{-\alpha} \mathbb{E}M_n \leq KN^{-\alpha}. \quad (\text{A.8})$$

Applying (A.1) and (A.8), we obtain

$$\mathbb{P}(F_2) \leq \mathbb{P}(M_n > N^\alpha) \leq KN^{-\alpha}. \quad (\text{A.9})$$

To estimate $\mathbb{P}(F_3)$, we observe that for N sufficiently large, each particle branches at rate at most σ . On F_1^c , it takes at most N^α units of time to reach the ultimate ancestor. Furthermore, on the set F_2^c , there are always at most N^α particles present, which implies that the probability of a collision per particle per branching event is at most N^α/N . Hence, the expected total number of branching events by time T_{UA}^N is bounded by $\sigma N^\alpha N^\alpha$. Combining these observations, we obtain

$$\mathbb{P}(F_3) \leq \sigma N^\alpha N^\alpha \frac{N^\alpha}{N} = \sigma N^{-(1-3\alpha)}. \quad (\text{A.10})$$

Finally, to estimate $\mathbb{P}(F_4)$, we need to look at the difference in the branching rates for $\mathcal{G}_n^N(t)$ and $\mathcal{G}_n(t)$; this is $|\sigma - Ns_N| k/2$ if both processes are of size k . On the set $F_1^c \cap F_2^c$, this rate is always at most $|\sigma - Ns_N| N^\alpha/2$. Furthermore, it will take at most N^α units of time until both processes reach their ultimate ancestors. On $F_1^c \cap F_2^c \cap F_3^c$, there are no fictitious particles and this allows

us to couple the two corresponding percolation diagrams. We thus obtain

$$\mathbb{P}(F_4) \leq \frac{1}{2} |\sigma - Ns_N| N^\alpha N^\alpha \leq \frac{1}{2} CN^{2\alpha-\gamma} \quad (\text{A.11})$$

where the last inequality follows from (1.5).

Combining (A.7), (A.9), (A.10) and (A.11), we get

$$\mathbb{P}(F_1 \cup F_2 \cup F_3 \cup F_4) \\ \leq 2KN^{-\alpha} + \sigma N^{-(1-3\alpha)} + \frac{1}{2} CN^{2\alpha-\gamma}. \quad (\text{A.12})$$

Since $\alpha < (\gamma/2) \wedge (1/3)$, the right hand side of (A.12) tends to 0 as $N \rightarrow \infty$. Observing that on $(F_1 \cup F_2 \cup F_3 \cup F_4)^c$ the two processes $\mathcal{G}_n^N(t)$ and $\mathcal{G}_n(t)$ agree for all $t \in [0, T_{UA}^N \vee T_{UA}]$, it thus follows that $\mathcal{G}_n^N(t)$ converges in distribution to $\mathcal{G}_n(t)$ as $N \rightarrow \infty$.

We still need to show that the mutation process converges to a point process in which mutation events occur according to the points of a Poisson process with rate $\theta/2$, independently along each branch of the graph. The first step is to show that mutation events do not occur at points where either a branching event or a coalescing event occurs. On the set $F_1^c \cap F_2^c$, the expected total number of branching or coalescing events by time $T_{UA}^N \vee T_{UA}$ is bounded by $(\sigma N^\alpha + N^{2\alpha})N^\alpha$. For each branching/coalescing event, there is probability u_N of a mutation. Set

$$F_5 = \{\text{a mutation event occurs at some coalescing} \\ \text{or branching point by time } T_{UA}^N \vee T_{UA}\} \\ \cap F_1^c \cap F_2^c \cap F_3^c \cap F_4^c.$$

Hence,

$$\mathbb{P}(F_5) \leq (\sigma N^\alpha + N^{2\alpha}) N^\alpha u_N \\ \leq (\sigma + 1) N^{3\alpha} \frac{Nu_N}{N} \\ \leq 2\theta(\sigma + 1) N^{-(1-3\alpha)}, \quad (\text{A.13})$$

for N sufficiently large.

To show that the mutation process $\mathcal{Y}_n^N(t)$ converges in distribution to $\mathcal{Y}_n(t)$, we estimate the difference. To do this, we set

$$F_6 = \{\text{a mutation event occurs in exactly} \\ \text{one of the two processes} \\ \mathcal{G}_n^N(t) \text{ and } \mathcal{G}_n(t) \text{ at some } t \leq T_{UA}^N \vee T_{UA}\} \\ \cap F_1^c \cap F_2^c \cap F_3^c \cap F_4^c.$$

An argument similar to the one used for the bound on $\mathbb{P}(F_4)$ yields

$$\mathbb{P}(F_6) \leq |Nu_N - \theta| N^{2\alpha} \leq CN^{2\alpha-\gamma}, \quad (\text{A.14})$$

where we again used (1.5). Since $\alpha < (\gamma/2) \wedge (1/3)$, the right hand side of (A.14) tends to 0 as $N \rightarrow \infty$.

The independence property follows immediately from the construction of the biased voter model since non-overlapping parts of the percolation diagram are independent. Combining this observation with (A.12), (A.13) and (A.14) shows that the mutation process $\mathcal{Y}_n^N(t)$ converges in distribution to $\mathcal{Y}_n(t)$ and has the properties described in the definition of $\mathcal{Y}_n(t)$. ■

APPENDIX B: IDENTITY BY DESCENT

We will continue here the (rather lengthy) calculation for finding the probability of identity by descent for the graphs in Fig. 7. In graph $g(1, 1)$ we may allow an even number of mutations in the dotted segment. This yields

$$\begin{aligned} & \mathbb{P}(\text{identity by descent, } \tilde{G} = g(1, 1) \mid G = g(1)) \\ &= \frac{1+\sigma}{1+\sigma+\theta} \left(\int_0^\infty \left(3 + \frac{3\sigma}{2}\right) e^{-\theta t} \frac{1+e^{-\theta t}}{2} e^{-(3+3\sigma/2)t} dt \right) \\ & \quad \times \frac{1+\sigma}{1+\sigma+\theta} \mathbb{P}(UA=1) \end{aligned} \quad (\text{B.1})$$

Similarly, in graph $g(1, 3)$, if the ultimate ancestor is of type 2, we may allow an even number of mutations on the dotted segment and no mutations in the dashed segments. This yields

$$\begin{aligned} & \mathbb{P}(\text{identity by descent, } \tilde{G} = g(1, 3) \mid G = g(1)) \\ &= \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \cdot \frac{2+2\sigma+\theta}{2+2\sigma+2\theta} \cdot \mathbb{P}(UA=2) \end{aligned} \quad (\text{B.2})$$

In graph $g(3, 1)$, the ultimate ancestor is of type 1 and we may allow an even number of mutations in both of the two dotted segments. We need to split the larger of the two dotted segments into two parts at the time of the MRCA. Thus, we obtain an even number of mutations in

the dotted segment if its two parts contain either both an even or both an odd number of mutations. We present the details for $g(3, 1)$ where we use the same labelling as in Fig. 8:

$$\begin{aligned} & \mathbb{P}(\text{identity by descent, } \tilde{G} = g(3, 1) \mid G = g(3)) \\ &= \mathbb{E} e^{-\theta T(1)} \left\{ \left(\int_0^\infty \left(3 + \frac{3\sigma}{2}\right) e^{-\theta t} \frac{1+e^{-\theta t}}{2} e^{-(3+3\sigma/2)t} dt \right) \right. \\ & \quad \times \left(\int_0^\infty (1+\sigma) \left(\frac{1+e^{-\theta t}}{2} \right)^2 e^{-(1+\sigma)t} dt \right) \\ & \quad \left. + \left(\int_0^\infty \left(3 + \frac{3\sigma}{2}\right) e^{-\theta t} \frac{1-e^{-\theta t}}{2} e^{-(3+3\sigma/2)t} dt \right) \right. \\ & \quad \left. \times \left(\int_0^\infty (1+\sigma) \frac{1+e^{-\theta t}}{2} \frac{1-e^{-\theta t}}{2} e^{-(1+\sigma)t} dt \right) \right\} \\ & \quad \times \mathbb{P}(UA=1) \\ &= \frac{1+\sigma}{1+\sigma+\theta} \left\{ \frac{9 \left(1 + \frac{\sigma}{2}\right) (2+\sigma+\theta)}{2 \left(3 + \frac{3\sigma}{2} + \theta\right) \left(3 + \frac{3\sigma}{2} + 2\theta\right)} \right. \\ & \quad \times \frac{2+4\sigma+2\sigma^2+4\theta+4\sigma\theta+\theta^2}{2(1+\sigma+\theta)(1+\sigma+2\theta)} \\ & \quad \left. + \frac{3 \left(1 + \frac{\sigma}{2}\right) \theta}{2 \left(3 + \frac{3\sigma}{2} + \theta\right) \left(3 + \frac{3\sigma}{2} + \sigma\right)} \cdot \frac{\theta}{2+2\sigma+4\theta} \right\} \\ & \quad \times \mathbb{P}(UA=1) \\ &= \frac{1+\sigma}{1+\sigma+\theta} 3 \left(1 + \frac{\sigma}{2}\right) \mathbb{P}(UA=1) \\ & \quad \times \frac{\left(3(2+\sigma+\theta)(2+4\sigma+2\sigma^2+4\theta+4\sigma\theta+\theta^2) + \theta^2(1+\sigma+\theta)\right)}{4 \left(3 + \frac{3\sigma}{2} + \theta\right) \left(3 + \frac{3\sigma}{2} + 2\theta\right) (1+\sigma+\theta)(1+\sigma+2\theta)}. \end{aligned} \quad (\text{B.3})$$

In graph $g(3, 2)$, the ultimate ancestor is of type 2 and we may allow any number of mutations in the unlabelled segment but again no mutations in the labelled segments. Graph $g(3, 3)$ is similar to $g(3, 1)$. In graph $g(3, 3)$, the ultimate ancestor is of type 2, the shorter of the two dotted segments must have an even number of mutations, the longer of the two dotted segments must have an odd number of mutations. Again, we need to split the longer of

the two segments into two parts. In graph $g(5, 1)$, if the ultimate ancestor is of type 1, we must have an even number of mutations in the dotted segment. This is basically the same calculation as for graph $g(1, 1)$. In graph $g(5, 2)$, the ultimate ancestor is of type 2 and therefore, we may have any number of mutations in the unlabelled segment of the graph. Because of symmetry, the calculations for $g(2)$ are identical to the one for $g(1)$, similarly with the other two pairs of symmetric graphs. This gives a factor of 2 in the calculation below in (B.4). We omit the calculations for the last four of the graphs in Fig. 7 and present the final result where the terms appear in the same order as in Fig. 7:

$$\begin{aligned}
& \sum_{g \in \mathcal{G}(2,3,2,1)} \mathbb{P}(\text{identity by descent} \mid G = g) P(G = g) \\
&= 2 \cdot \frac{1}{6} \cdot \frac{\sigma}{1+\sigma} \cdot \frac{2}{2+\sigma} \cdot \frac{1}{1+\sigma} \\
& \times \left\{ \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{9 \left(1 + \frac{\sigma}{2}\right) (2+\sigma+\theta)}{2 \left(3 + \frac{3\sigma}{2} + \theta\right) \left(3 + \frac{3\sigma}{2} + 2\theta\right)} \right. \\
& \times \frac{1+\sigma}{1+\sigma+\theta} \cdot \mathbb{P}(UA = 1) \\
& + \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \cdot \frac{\theta}{2 + 2\sigma + 2\theta} \cdot \mathbb{P}(UA = 1) \\
& + \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \cdot \frac{2 + 2\sigma + \theta}{2 + 2\sigma + 2\theta} \cdot \mathbb{P}(UA = 2) \\
& + \frac{1+\sigma}{1+\sigma+\theta} 3 \left(1 + \frac{\sigma}{2}\right) \mathbb{P}(UA = 1) \\
& \times \frac{\left(3(2+\sigma+\theta)(2+4\sigma+2\sigma^2+4\theta+4\sigma\theta+\theta^2)\right)}{\theta^2(1+\sigma+\theta)} \\
& \left. \times \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \cdot \frac{1+\sigma}{1+\sigma+\theta} \cdot \mathbb{P}(UA = 2) \right\} \\
& + \frac{1+\sigma}{1+\sigma+\theta} 3 \left(- + \frac{\sigma}{2}\right) \mathbb{P}(UA = 2)
\end{aligned}$$

$$\begin{aligned}
& \times \frac{\left(3(2+\sigma+\theta)\theta(1+\sigma+\theta)\right)}{\theta(2+4\sigma+2\sigma^2+4\theta+4\sigma\theta+\theta^2)} \\
& \times \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{9 \left(1 + \frac{\sigma}{2}\right) (2+\sigma+\theta)}{2 \left(3 + \frac{3\sigma}{2} + \theta\right) \left(3 + \frac{3\sigma}{2} + 2\theta\right)} \\
& \times \frac{1+\sigma}{1+\sigma+\theta} \cdot \mathbb{P}(UA = 1) \\
& + \frac{1+\sigma}{1+\sigma+\theta} \cdot \frac{3 + \frac{3\sigma}{2}}{3 + \frac{3\sigma}{2} + \theta} \cdot \frac{1+\sigma}{1+\sigma+\theta} \cdot \mathbb{P}(UA = 2) \left. \right\}. \quad (\text{B.4})
\end{aligned}$$

APPENDIX C: LAPLACE TRANSFORM OF T_2

In this part of the appendix, we will find an expression for the Laplace transform of T_2 which we can use to find the power series expansion for small σ . We begin with evaluating (4.23) and (4.24). Because of the independence of mutations along disjoint pieces of the ancestral selection graph, we have

$$\begin{aligned}
& \mathbb{P}(M(T(3) + T(2)) \text{ is even}) \\
&= \mathbb{P}(M(T(3)) \text{ is even}) \mathbb{P}(M(T(2)) \text{ is even}) \\
& \quad + \mathbb{P}(M(T(3)) \text{ is odd}) \mathbb{P}(M(T(2)) \text{ is odd}) \\
& \times \left(\int_0^\infty (1+\sigma) e^{-(1+\sigma)t} \frac{1+e^{-\theta t}}{2} dt \right) \\
& \times \left(\int_0^\infty (3+3\sigma/2) e^{-(3+3\sigma/2)s} \frac{1+e^{-\theta s}}{2} ds \right) \\
& + \left(\int_0^\infty (1+\sigma) e^{-(1+\sigma)t} \frac{1-e^{-\theta t}}{2} dt \right) \\
& \times \left(\int_0^\infty (3+3\sigma/2) e^{-(3+3\sigma/2)s} \frac{1-e^{-\theta s}}{2} ds \right) \\
&= \left(\frac{1}{2} + \frac{1}{2} \frac{1+\sigma}{1+\sigma+\theta} \right) \left(\frac{1}{2} + \frac{1}{2} \frac{3+3\sigma/2}{3+3\sigma/2+\theta} \right) \\
& \quad + \left(\frac{1}{2} - \frac{1}{2} \frac{1+\sigma}{1+\sigma+\theta} \right) \left(\frac{1}{2} - \frac{1}{2} \frac{3+3\sigma/2}{3+3\sigma/2+\theta} \right) \\
&= \frac{1}{2} + \frac{1}{2} \left(\frac{1+\sigma}{1+\sigma+\theta} \right) \left(\frac{3+3\sigma/2}{3+3\sigma/2+\theta} \right). \quad (\text{C.1})
\end{aligned}$$

Similarly,

$$\begin{aligned} & \mathbb{P}(M(T(3) + T(2)) \text{ is odd}) \\ &= \mathbb{P}(M(T(3)) \text{ is even}) \mathbb{P}(M(T(2)) \text{ is odd}) \\ & \quad + \mathbb{P}(M(T(3)) \text{ is odd}) \mathbb{P}(M(T(2)) \text{ is even}) \\ &= \frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right). \end{aligned} \quad (\text{C.2})$$

Combining (C.1) and (C.2), we have

$$\begin{aligned} & \mathbb{E}[e^{-\alpha T_2} \mid G = g(1)] \\ &= \mathbb{E}[e^{-\alpha T_2} \mid T_2 = T(1) + T(2) + T(3), G = g(1)] \\ & \quad \times \mathbb{P}(T_2 = T(1) + T(2) + T(3) \mid G = g(1)) \\ & \quad + \mathbb{E}[e^{-\alpha T_2} \mid T_2 = T(1) + T(2), G = g(1)] \\ & \quad \times \mathbb{P}(T_2 = T(1) + T(2) \mid G = g(1)) \\ &= \mathbb{E}[e^{-\alpha T(1)}] \mathbb{E}[e^{-\alpha T(2)}] \mathbb{E}[e^{-\alpha T(3)}] \\ & \quad \times \left(\mathbb{P}(UA = 1) \left[\frac{1}{2} + \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right. \\ & \quad \left. + \mathbb{P}(UA = 2) \left[\frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right) \\ & \quad + \mathbb{E}[e^{-\alpha T(1)}] \mathbb{E}[e^{-\alpha T(2)}] \\ & \quad \times \left(\mathbb{P}(UA = 1) \left[\frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right. \\ & \quad \left. + \mathbb{P}(UA = 2) \left[\frac{1}{2} + \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right) \\ &= \left(\frac{1 + \sigma}{1 + \sigma + \alpha} \right)^2 \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \alpha} \right) \\ & \quad \times \left(\mathbb{P}(UA = 1) \left[\frac{1}{2} + \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right. \\ & \quad \left. + \mathbb{P}(UA = 2) \left[\frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right) \\ & \quad + \left(\frac{1 + \sigma}{1 + \sigma + \alpha} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \alpha} \right) \\ & \quad \times \left(\mathbb{P}(UA = 1) \left[\frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right. \\ & \quad \left. + \mathbb{P}(UA = 2) \left[\frac{1}{2} + \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right). \end{aligned} \quad (\text{C.3})$$

By symmetry, $\mathbb{E}[e^{-\alpha T_2} \mid G = g(2)]$ has the same value.

Using the same type of argument, it is easy to see that

$$\begin{aligned} & \mathbb{E}[e^{-\alpha T_2} \mid G = g(3)] = \mathbb{E}[e^{-\alpha T_2} \mid G = g(4)] \\ &= \left(\frac{1 + \sigma}{1 + \sigma + \alpha} \right)^2 \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \alpha} \right) \\ & \quad \times \left(\mathbb{P}(UA = 1) \left[\frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right. \\ & \quad \left. + \mathbb{P}(UA = 2) \left[\frac{1}{2} + \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right) \\ & \quad + \left(\frac{1 + \sigma}{1 + \sigma + \alpha} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \alpha} \right) \\ & \quad \times \left(\mathbb{P}(UA = 1) \left[\frac{1}{2} + \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right. \\ & \quad \left. + \mathbb{P}(UA = 2) \left[\frac{1}{2} - \frac{1}{2} \left(\frac{1 + \sigma}{1 + \sigma + \theta} \right) \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \theta} \right) \right] \right). \end{aligned} \quad (\text{C.4})$$

Finally, for $g(5)$ and $g(6)$, there is only one possible MRCA, namely the UA. Thus,

$$\begin{aligned} & \mathbb{E}[e^{-\alpha T_2} \mid G = g(5)] = \mathbb{E}[e^{-\alpha T_2} \mid G = g(6)] \\ &= \left(\frac{1 + \sigma}{1 + \sigma + \alpha} \right)^2 \left(\frac{3 + 3\sigma/2}{3 + 3\sigma/2 + \alpha} \right). \end{aligned} \quad (\text{C.5})$$

ACKNOWLEDGMENTS

The authors thank Simon Tavaré for introducing us to the problem. We also thank two referees for their helpful comments.

REFERENCES

- Abramowitz, M., and Stegun, I. 1972. "Handbook of Mathematical Functions," revised ed., Dover, New York.
- Bramson, M., and Griffeath, D. 1980. On the Williams–Bjerknes tumor growth model, II, *Math. Proc. Cambridge Phil. Soc.* **88**, 339–357.
- Bramson, M., and Griffeath, D. 1981. On the Williams–Bjerknes tumor growth model, I, *Ann. Probab.* **9**, 173–185.
- Cannings, C. 1974. The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models, *Adv. Appl. Probab.* **6**, 260–290.
- Clifford, P., and Sudbury, A. 1973. A model for spatial conflict, *Biometrika* **60**, 581–588.
- Donnelly, P. 1984. The transient behaviour of the Moran model in population genetics, *Proc. Camb. Phil. Soc.* **95**, 349–358.

- Donnelly, P., and Taveré, S. 1997. "Ancestral Processes in Population Genetics," in preparation.
- Durrett, R., and Neuhauser, C. 1994. Particle systems and reaction-diffusion equations, *Ann. Probab.* **22**, 289–333.
- Ethier, S. N., and Kurtz, T. G. 1987. The infinitely-many-alleles model with selection as a measure-valued diffusion, in "Lecture Notes in Biomathematics," Vol. 70, pp. 72–86, Springer, New York.
- Ethier, S. N., and Kurtz, T. G. 1994. Convergence to Fleming–Viot processes in the weak atomic topology, *Stoch. Proc. Appl.* **54**, 1–27.
- Fisher, R. A. 1930. "The Genetical Theory of Natural Selection," Clarendon Press, Oxford.
- Golding, G. B. 1997. The effect of purifying selection on genealogies, in "Progress in Population Genetics and Human Evolution" (P. Donnelly and S. Tavaré, Eds.), IMA Proceedings, **87**, Springer-Verlag, Berlin/New York.
- Griffiths, R. C. 1989. Genealogical-tree probabilities in the infinitely-many-site model, *J. Math. Biol.* **27**, 667–680.
- Griffiths, R. C. 1991. The two-locus ancestral graph, in "Selected Proceedings of the Symposium on Applied Probability, Sheffield, 1989" (I. V. Basewa and R. L. Taylor, Eds.), pp. 100–117, IMS Lecture Notes—Monograph Series, Vol. 18, Institute of Mathematical Statistics.
- Griffiths, R. C., and Marjoram, P. 1997. An ancestral recombination graph, in "Progress in Population Genetics and Human Evolution" (P. Donnelly and S. Tavaré, Eds.), IMA Proceedings, **87**, Springer-Verlag, Berlin/New York.
- Griffiths, R. C., and Tavaré, S. 1994. Simulating probability distributions in the coalescent, *Theor. Population Biology* **46**, 131–159.
- Harris, T. E. 1972. Nearest neighbor Markov interaction processes on multidimensional lattices, *Adv. Math.* **9**, 66–89.
- Harris, T. E. 1976. On a class of set-valued Markov processes, *Ann. Probab.* **4**, 175–194.
- Holley, R., and Liggett, T. M. 1975. Ergodic theorems for weakly interacting systems and the voter model, *Ann. Probab.* **3**, 643–663.
- Hudson, R. R., and Kaplan, N. L. 1988. The coalescent process in models with selection and recombination, *Genetics* **120**, 831–840.
- Joyce, P., and Tavaré, S. 1995. The distribution of rare alleles, *J. Math. Biol.* **33**, 602–618.
- Kaplan, N. L., Darden, T., and Hudson, R. R. 1988. The coalescent process in models with selection, *Genetics* **120**, 819–829.
- Karlin, S., and Taylor, H. M. 1975. "A First Course in Stochastic Processes," 2nd ed., Academic Press, New York.
- Kimura, M. 1956. "Stochastic Processes in Population Genetics," Ph.D. thesis, University of Wisconsin, Madison.
- Kingman, J. F. C. 1982a. The coalescent, *Stoch. Proc. Appl.* **13**, 235–248.
- Kingman, J. F. C. 1982b. On the genealogy of large populations, *J. Appl. Prob.* **19A**, 27–43.
- Liggett, T. M. 1985. "Interacting Particle Systems," Springer, New York.
- Lundstrom, R. 1990. "Stochastic Models and Statistical Methods for DNA Sequence Data," Ph.D. thesis, Department of Mathematics, University of Utah.
- Moran, P. A. P. 1958. Random processes in genetics, *Proc. Cambridge Phil. Soc.* **54**, 60–71.
- Moran, P. A. P. 1962. "The Statistical Processes of Evolutionary Theory," Clarendon Press, Oxford.
- Neuhauser, C., and Krone, S. 1997. The genealogy of samples in models with selection, *Genetics* **154**, 519–534.
- Sawyer, S., Dykhuizen, D., and Hartl, D. 1987. Confidence interval for the number of selectively neutral amino acid polymorphisms, *Proc. Natl. Acad. Sci. USA* **84**, 6225–6228.
- Schwartz, D. L. 1977. Application of duality to a class of Markov processes, *Ann. Probab.* **5**, 522–532.
- Tavaré, S. 1984. Line-of-descent and genealogical processes, and their applications in population genetics models, *Theor. Popul. Biol.* **26**, 119–164.
- Watterson, G. A. 1977. Heterosis or neutrality?, *Genetics* **86**, 789–814.
- Watterson, G. A. 1978. The homozygosity test of neutrality, *Genetics* **88**, 405–417.
- Watterson, G. A. 1984. Lines of descent and the coalescent, *Theor. Popul. Biol.* **26**, 77–92.
- Wright, S. 1931. Evolution in Mendelian populations, *Genetics* **16**, 97–159.
- Wright, S. 1949. Adaptation and selection, in "Genetics, Paleontology and Evolution" (G. L. Jepson, G. G. Simpson, and E. Mayr, Eds.), pp. 365–389, Princeton Univ. Press, Princeton.